(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 October 2002 (24.10.2002)

PCT

(10) International Publication Number
WO 02/083873 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/15253

(22) International Filing Date: 10 April 2002 (10.04.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

| | | |
|---|---|---|
| 60/283,793 | 13 April 2001 (13.04.2001) | US |
| 60/291,544 | 16 May 2001 (16.05.2001) | US |
| 60/293,572 | 25 May 2001 (25.05.2001) | US |
| 60/308,182 | 27 July 2001 (27.07.2001) | US |
| 60/311,447 | 9 August 2001 (09.08.2001) | US |
| 60/315,874 | 29 August 2001 (29.08.2001) | US |
| 60/322,181 | 14 September 2001 (14.09.2001) | US |

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **SANJANWALA, Madhusudan, M.** [US/US]; 210 Silvia Court, Los Altos, CA 94024 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). **DING, Li** [CN/US]; 3353 Alma Street # 146, Palo Alto, CA 94306 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **LI, Joana, X.** [US/US]; 1264 Geneva Avenue, San Francisco, CA 94112 (US). **LAL, Preeti, G.** [US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **LU, Dyung Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **GORVAD, Ann, E.** [US/US]; 369 Marie Common, Livermore, CA 94550 (US). **FORSYTHE, Ian, J.** [US/US]; 308 Roble Avenue, Redwood City, CA 94061 (US). **DUGGAN, Brendan, M.** [AU/US]; 243 Buena Vista Avenue # 306, Sunnyvale, CA 94086 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue #23, Mountain View, CA 94043 (US). **EMERLING, Brooke, M.** [US/US]; 1735 Woodland Avenue # 71, Palo Alto, CA 94303 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **BECHA, Shanya** [US/US]; 21062 Gary Drive # 117, Castro Valley, CA 94546 (US). **SPRAGUE, William, W.** [US/US]; 611 13th Street # C, Sacramento, CA 95814 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— *without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: ENZYMES

(57) Abstract: The invention provides human enzymes (ENZM) and polynucleotides which identify and encode ENZM. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of ENZM.



WO 02/083873 A2

# ENZYMES

## TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of enzymes and to the use of these sequences in the diagnosis, treatment, and prevention of autoimmune/inflammatory disorders, infectious disorders, immune deficiencies, disorders of metabolism, reproductive disorders, neurological disorders, cardiovascular disorders, eye disorders, and cell proliferative disorders, including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of enzymes.

## BACKGROUND OF THE INVENTION

The cellular processes of biogenesis and biodegradation involve a number of key enzyme classes including oxidoreductases, transferases, hydrolases, lyases, isomerases, ligases, and others. Each class of enzyme comprises many substrate-specific enzymes having precise and well regulated functions. Enzymes facilitate metabolic processes such as glycolysis, the tricarboxylic cycle, and fatty acid metabolism; synthesis or degradation of amino acids, steroids, phospholipids, and alcohols; regulation of cell signaling, proliferation, inflamation, and apoptosis; and through catalyzing critical steps in DNA replication and repair and the process of translation.

Oxidoreductases

Many pathways of biogenesis and biodegradation require oxidoreductase (dehydrogenase or reductase) activity, coupled to reduction or oxidation of a cofactor. Potential cofactors include cytochromes, oxygen, disulfide, iron-sulfur proteins, flavin adenine dinucleotide (FAD), and the nicotinamide adenine dinucleotides NAD and NADP (Newsholme, E.A. and Leech, A.R. (1983) Biochemistry for the Medical Sciences, John Wiley and Sons, Chichester, U. K. pp. 779-793). Reductase activity catalyzes transfer of electrons between substrate(s) and cofactor(s) with concurrent oxidation of the cofactor. Reverse dehydrogenase activity catalyzes the reduction of a cofactor and consequent oxidation of the substrate. Oxidoreductase enzymes are a broad superfamily that catalyze reactions in all cells of organisms, including metabolism of sugar, certain detoxification reactions, and synthesis or degradation of fatty acids, amino acids, glucocorticoids, estrogens, androgens, and prostaglandins. Different family members may be referred to as oxidoreductases, oxidases, reductases, or dehydrogenases, and they often have distinct cellular locations such as the cytosol, the plasma membrane, mitochondrial inner or outer membrane, and peroxisomes.

Short-chain alcohol dehydrogenases (SCADs) are a family of dehydrogenases that share only 15% to 30% sequence identity, with similarity predominantly in the coenzyme binding domain and the substrate binding domain. In addition to their role in detoxification of ethanol, SCADs are

involved in synthesis and degradation of fatty acids, steroids, and some prostaglandins, and are therefore implicated in a variety of disorders such as lipid storage disease, myopathy, SCAD deficiency, and certain genetic disorders. For example, retinol dehydrogenase is a SCAD-family member (Simon, A. et al. (1995) J. Biol. Chem. 270:1107-1112) that converts retinol to retinal, the precursor of retinoic acid. Retinoic acid, a regulator of differentiation and apoptosis, has been shown to down-regulate genes involved in cell proliferation and inflammation (Chai, X. et al. (1995) J. Biol. Chem. 270:3900-3904). In addition, retinol dehydrogenase has been linked to hereditary eye diseases such as autosomal recessive childhood-onset severe retinal dystrophy (Simon, A. et al. (1996) Genomics 36:424-430).

Membrane-bound succinate dehydrogenases (succinate:quinone reductases, SQR) and fumarate reductases (quinol:fumarate reductases, QFR) couple the oxidation of succinate to fumarate with the reduction of quinone to quinol, and also catalyze the reverse reaction. QFR and SQR complexes are collectively known as succinate:quinone oxidoreductases (EC 1.3.5.1) and have similar compositions. The complexes consist of two hydrophilic and one or two hydrophobic, membrane-integrated subunits. The larger hydrophilic subunit A carries covalently bound flavin adenine dinucleotide; subunit B contains three iron-sulphur centers (Lancaster, C.R. and Kroger, A. (2000) Biochim. Biophys. Acta 1459:422-431). The full-length cDNA sequence for the flavoprotein subunit of human heart succinate dehydrogenase (succinate: (acceptor) oxidoreductase; EC 1.3.99.1) is similar to the bovine succinate dehydrogenase in that it contains a cysteine triplet and in that the active site contains an additional cysteine that is not present in yeast or prokaryotic SQRs (Morris, A. A. et al. (1994) Biochim. Biophys. Acta 29:125-128).

Propagation of nerve impulses, modulation of cell proliferation and differentiation, induction of the immune response, and tissue homeostasis involve neurotransmitter metabolism (Weiss, B. (1991) Neurotoxicology 12:379-386; Collins, S.M. et al. (1992) Ann. N.Y. Acad. Sci. 664:415-424; Brown, J.K. and Imam, H. (1991) J. Inherit. Metab. Dis. 14:436-458). Many pathways of neurotransmitter metabolism require oxidoreductase activity, coupled to reduction or oxidation of a cofactor, such as $NAD^+$/NADH (Newsholme, E.A. and Leech, A.R. (1983) Biochemistry for the Medical Sciences, John Wiley and Sons, Chichester, U.K. pp. 779-793). Degradation of catecholamines (epinephrine or norepinephrine) requires alcohol dehydrogenase (in the brain) or aldehyde dehydrogenase (in peripheral tissue). $NAD^+$-dependent aldehyde dehydrogenase oxidizes 5-hydroxyindole-3-acetate (the product of 5-hydroxytryptamine (serotonin) metabolism) in the brain, blood platelets, liver and pulmonary endothelium (Newsholme, E.A. and Leech, A.R. supra, p. 786). Other neurotransmitter degradation pathways that utilize $NAD^+$/NADH-dependent oxidoreductase activity include those of L-DOPA (precursor of dopamine, a neuronal excitatory compound), glycine (an inhibitory neurotransmitter in the brain and spinal cord), histamine (liberated from mast cells

during the inflammatory response), and taurine (an inhibitory neurotransmitter of the brain stem, spinal cord and retina) (Newsholme, E.A. and Leech, A.R. supra, pp. 790, 792). Epigenetic or genetic defects in neurotransmitter metabolic pathways can result in diseases including Parkinson disease and inherited myoclonus (McCance, K.L. and Huether, S.E. (1994) Pathophysiology, Mosby-Year Book, Inc., St. Louis, MO pp. 402-404; Gundlach, A.L. (1990) FASEB J. 4:2761-2766).

Tetrahydrofolate is a derivatized glutamate molecule that acts as a carrier, providing activated one-carbon units to a wide variety of biosynthetic reactions, including synthesis of purines, pyrimidines, and the amino acid methionine. Tetrahydrofolate is generated by the activity of a holoenzyme complex called tetrahydrofolate synthase, which includes three enzyme activities: tetrahydrofolate dehydrogenase, tetrahydrofolate cyclohydrolase, and tetrahydrofolate synthetase. Thus, tetrahydrofolate dehydrogenase plays an important role in generating building blocks for nucleic and amino acids, crucial to proliferating cells.

3-Hydroxyacyl-CoA dehydrogenase (3HACD) is involved in fatty acid metabolism. It catalyzes the reduction of 3-hydroxyacyl-CoA to 3-oxoacyl-CoA, with concomitant oxidation of NAD to NADH, in the mitochondria and peroxisomes of eukaryotic cells. In peroxisomes, 3HACD and enoyl-CoA hydratase form an enzyme complex called bifunctional enzyme, defects in which are associated with peroxisomal bifunctional enzyme deficiency. This interruption in fatty acid metabolism produces accumulation of very-long chain fatty acids, disrupting development of the brain, bone, and adrenal glands. Infants born with this deficiency typically die within 6 months (Watkins, P. et al. (1989) J. Clin. Invest. 83:771-1277; Online Mendelian Inheritance in Man (OMIM), #261515). The neurodegeneration characteristic of Alzheimer's disease involves development of extracellular plaques in certain brain regions. A major protein component of these plaques is the peptide amyloid-β (Aβ), which is one of several cleavage products of amyloid precursor protein (APP). 3HACD has been shown to bind the Aβ peptide, and is overexpressed in neurons affected in Alzheimer's disease. In addition, an antibody against 3HACD can block the toxic effects of Aβ in a cell culture model of Alzheimer's disease (Yan, S. et al. (1997) Nature 389:689-695; OMIM, #602057).

Steroids such as estrogen, testosterone, and corticosterone are generated from a common precursor, cholesterol, and interconverted. Enzymes acting upon cholesterol include dehydrogenases. Steroid dehydrogenases, such as the hydroxysteroid dehydrogenases, are involved in hypertension, fertility, and cancer (Duax, W.L. and Ghosh, D. (1997) Steroids 62:95-100). One such dehydrogenase is 3-oxo-5-α-steroid dehydrogenase (OASD), a microsomal membrane protein highly expressed in prostate and other androgen-responsive tissues. OASD catalyzes the conversion of testosterone into dihydrotestosterone, which is the most potent androgen. Dihydrotestosterone is essential for the formation of the male phenotype during embryogenesis, as well as for proper

androgen-mediated growth of tissues such as the prostate and male genitalia. A defect in OASD leads to defective formation of the external genitalia (Andersson, S. et al. (1991) Nature 354:159-161; Labrie, F. et al. (1992) Endocrinology 131:1571-1573; OMIM #264600).

17β-hydroxysteroid dehydrogenase (17βHSD6) plays an important role in the regulation of the male reproductive hormone, dihydrotestosterone (DHTT). 17βHSD6 acts to reduce levels of DHTT by oxidizing a precursor of DHTT, 3α-diol, to androsterone which is readily glucuronidated and removed. 17βHSD6 is active with both androgen and estrogen substrates in embryonic kidney 293 cells. Isozymes of 17βHSD catalyze oxidation and/or reduction reactions in various tissues with preferences for different steroid substrates (Biswas, M.G. and Russell, D.W. (1997) J. Biol. Chem. 272:15959-15966). For example, 17βHSD1 preferentially reduces estradiol and is abundant in the ovary and placenta. 17βHSD2 catalyzes oxidation of androgens and is present in the endometrium and placenta. 17βHSD3 is exclusively a reductive enzyme in the testis (Geissler, W.M. et al. (1994) Nature Genet. 7:34-39). An excess of androgens such as DHTT can contribute to diseases such as benign prostatic hyperplasia and prostate cancer.

The oxidoreductase isocitrate dehydrogenase catalyzes the conversion of isocitrate to α-ketoglutarate, a substrate of the citric acid cycle. Isocitrate dehydrogenase can be either NAD or NADP dependent, and is found in the cytosol, mitochondria, and peroxisomes. Activity of isocitrate dehydrogenase is regulated developmentally, and by hormones, neurotransmitters, and growth factors.

Hydroxypyruvate reductase (HPR), a peroxisomal 2-hydroxyacid dehydrogenase in the glycolate pathway, catalyzes the conversion of hydroxypyruvate to glycerate with the oxidation of both NADH and NADPH. The reverse dehydrogenase reaction reduces $NAD^+$ and $NADP^+$. HPR recycles nucleotides and bases back into pathways leading to the synthesis of ATP and GTP, which are used to produce DNA and RNA and to control various aspects of signal transduction and energy metabolism. Purine nucleotide biosynthesis inhibitors are used as antiproliferative agents to treat cancer and viral diseases. HPR also regulates biochemical synthesis of serine and cellular serine levels available for protein synthesis.

The mitochondrial electron transport (or respiratory) chain is the series of oxidoreductase-type enzyme complexes in the mitochondrial membrane that is responsible for the transport of electrons from NADH to oxygen and the coupling of this oxidation to the synthesis of ATP (oxidative phosphorylation). ATP provides energy to drive energy-requiring reactions. The key respiratory chain complexes are NADH:ubiquinone oxidoreductase (complex I), succinate:ubiquinone oxidoreductase (complex II), cytochrome $c_1$-b oxidoreductase (complex III), cytochrome c oxidase (complex IV), and ATP synthase (complex V) (Alberts, B. et al. (1994) Molecular Biology of the Cell, Garland Publishing, Inc., New York, NY, pp. 677-678). All of these complexes are located on

the inner matrix side of the mitochondrial membrane except complex II, which is on the cytosolic side where it transports electrons generated in the citric acid cycle to the respiratory chain. Electrons released in oxidation of succinate to fumarate in the citric acid cycle are transferred through electron carriers in complex II to membrane bound ubiquinone (Q). Transcriptional regulation of these nuclear-encoded genes controls the biogenesis of respiratory enzymes. Defects and altered expression of enzymes in the respiratory chain are associated with a variety of disease conditions.

Other dehydrogenase activities using NAD as a cofactor include 3-hydroxyisobutyrate dehydrogenase (3HBD), which catalyzes the NAD-dependent oxidation of 3-hydroxyisobutyrate to methylmalonate semialdehyde within mitochondria. 3-hydroxyisobutyrate levels are elevated in ketoacidosis, methylmalonic acidemia, and other disorders (Rougraff, P.M. et al. (1989) J. Biol. Chem. 264:5899-5903). Another mitochondrial dehydrogenase important in amino acid metabolism is the enzyme isovaleryl-CoA-dehydrogenase (IVD). IVD is involved in leucine metabolism and catalyzes the oxidation of isovaleryl-CoA to 3-methylcrotonyl-CoA. Human IVD is a tetrameric flavoprotein synthesized in the cytosol with a mitochondrial import signal sequence. A mutation in the gene encoding IVD results in isovaleric acidemia (Vockley, J. et al. (1992) J. Biol. Chem. 267:2494-2501).

The family of glutathione peroxidases encompass tetrameric glutathione peroxidases (GPx1-3) and the monomeric phospholipid hydroperoxide glutathione peroxidase (PHGPx/GPx4). Although the overall homology between the tetrameric enzymes and GPx4 is less than 30%, a pronounced similarity has been detected in clusters involved in the active site and a common catalytic triad has been defined by structural and kinetic data (Epp, O. et al. (1983) Eur. J. Biochem. 133:51-69). GPx1 is ubiquitously expressed in cells, whereas GPx2 is present in the liver and colon, and GPx3 is present in plasma. GPx4 is found at low levels in all tissues but is expressed at high levels in the testis (Ursini, F. et al (1995) Meth. Enzymol. 252:38-53). GPx4 is the only monomeric glutathione peroxidase found in mammals and the only mammalian glutathione peroxidase to show high affinity for and reactivity with phospholipid hydroperoxides, and to be membrane associated. A tandem mechanism for the antioxidant activities of GPx4 and vitamin E has been suggested. GPx4 has alternative transcription and translation start sites which determine its subcellular localization (Esworthy, R.S. et al. (1994) Gene 144:317-318; and Maiorino, M. et al. (1990) Meth. Enzymol. 186:448-450).

The glutathione S-transferases (GST) are a ubiquitous family of enzymes with dual substrate specificities that perform important biochemical functions of xenobiotic biotransformation and detoxification, drug metabolism, and protection of tissues against peroxidative damage. They catalyze the conjugation of an electrophile with reduced glutathione (GSH) which results in either activation or deactivation/detoxification. The absolute requirement for binding reduced GSH to a

variety of chemicals necessitates a diversity in GST structures in various organisms and cell types. GSTs are homodimeric or heterodimeric proteins localized in the cytosol. The major isozymes share common structural and catalytic properties and include four major classes, Alpha, Mu, Pi, and Theta. Each GST possesses a common binding site for GSH, and a variable hydrophobic binding site specific for its particular electrophilic substrates. Specific amino acid residues within GSTs have been identified as important for these binding sites and for catalytic activity. Residues Q67, T68, D101, E104, and R131 are important for the binding of GSH (Lee, H.-C. et al. (1995) J. Biol. Chem. 270:99-109). Residues R13, R20, and R69 are important for the catalytic activity of GST (Stenberg, G. et al. (1991) Biochem. J. 274:549-555).

GSTs normally deactivate and detoxify potentially mutagenic and carcinogenic chemicals. Some forms of rat and human GSTs are reliable preneoplastic markers of carcinogenesis. Dihalomethanes, which produce liver tumors in mice, are believed to be activated by GST (Thier, R. et al. (1993) Proc. Natl. Acad. Sci. USA 90:8567-8580). The mutagenicity of ethylene dibromide and ethylene dichloride is increased in bacterial cells expressing the human Alpha GST, A1-1, while the mutagenicity of aflatoxin B1 is substantially reduced by enhancing the expression of GST (Simula, T.P. et al. (1993) Carcinogenesis 14:1371-1376). Thus, control of GST activity may be useful in the control of mutagenesis and carcinogenesis.

GST has been implicated in the acquired resistance of many cancers to drug treatment, the phenomenon known as multi-drug resistance (MDR). MDR occurs when a cancer patient is treated with a cytotoxic drug such as cyclophosphamide and subsequently becomes resistant to this drug and to a variety of other cytotoxic agents as well. Increased GST levels are associated with some drug resistant cancers, and it is believed that this increase occurs in response to the drug agent which is then deactivated by the GST catalyzed GSH conjugation reaction. The increased GST levels then protect the cancer cells from other cytotoxic agents for which GST has affinity. Increased levels of A1-1 in tumors has been linked to drug resistance induced by cyclophosphamide treatment (Dirven, H.A. et al. (1994) Cancer Res. 54:6215-6220). Thus control of GST activity in cancerous tissues may be useful in treating MDR in cancer patients.

The reduction of ribonucleotides to the corresponding deoxyribonucleotides, needed for DNA synthesis during cell proliferation, is catalyzed by the enzyme ribonucleotide diphosphate reductase. Glutaredoxin is a glutathione (GSH)-dependent hydrogen donor for ribonucleotide diphosphate reductase and contains the active site consensus sequence -C-P-Y-C-. This sequence is conserved in glutaredoxins from such different organisms as E. coli, vaccinia virus, yeast, plants, and mammalian cells. Glutaredoxin has inherent GSH-disulfide oxidoreductase (thioltransferase) activity in a coupled system with GSH, NADPH, and GSH-reductase, catalyzing the reduction of low molecular weight disulfides as well as proteins. Glutaredoxin has been proposed to exert a general thiol redox

control of protein activity by acting both as an effective protein disulfide reductase, similar to thioredoxin, and as a specific GSH-mixed disulfide reductase (Padilla, C.A. et al. (1996) FEBS Lett. 378:69-73).

In addition to their important role in DNA synthesis and cell division, glutaredoxin and other thioproteins provide effective antioxidant defense against oxygen radicals and hydrogen peroxide (Schallreuter, K.U. and J.M. Wood (1991) Melanoma Res. 1:159-167). Glutaredoxin is the principal agent responsible for protein dethiolation in vivo and reduces dehydroascorbic acid in normal human neutrophils (Jung, C.H. and J.A. Thomas (1996) Arch. Biochem. Biophys. 335:61-122; Park, J.B. and M. Levine (1996) Biochem. J. 315:931-938).

The thioredoxin system serves as a hydrogen donor for ribonucleotide reductase and as a regulator of enzymes by redox control. It also modulates the activity of transcription factors such as NF-κB, AP-1, and steroid receptors. Several cytokines or secreted cytokine-like factors such as adult T-cell leukemia-derived factor, 3B6-interleukin-1, T-hybridoma-derived (MP-6) B cell stimulatory factor, and early pregnancy factor have been reported to be identical to thioredoxin (Holmgren, A. (1985) Annu. Rev. Biochem. 54:237-271; Abate, C. et al. (1990) Science 249:1157-1161; Tagaya, Y. et al. (1989) EMBO J. 8:757-764; Wakasugi, H. (1987) Proc. Natl. Acad. Sci. USA 84:804-808; Rosen, A. et al. (1995) Int. Immunol. 7:625-633). Thus thioredoxin secreted by stimulated lymphocytes (Yodoi, J. and T. Tursz (1991) Adv. Cancer Res. 57:381-411; Tagaya, N. et al. (1990) Proc. Natl. Acad. Sci. USA 87:8282-8286) has extracellular activities including a role as a regulator of cell growth and a mediator in the immune system (Miranda-Vizuete, A. et al. (1996) J. Biol. Chem. 271:19099-19103; Yamauchi, A. et al. (1992) Mol. Immunol. 29:263-270). Thioredoxin and thioredoxin reductase protect against cytotoxicity mediated by reactive oxygen species in disorders such as Alzheimer's disease (Lovell, M.A. (2000) Free Radic. Biol. Med. 28:418-427).

The selenoprotein thioredoxin reductase is secreted by both normal and neoplastic cells and has been implicated as both a growth factor and as a polypeptide involved in apoptosis (Soderberg, A. et al. (2000) Cancer Res. 60:2281-2289). An extracellular plasmin reductase secreted by hamster ovary cells (HT-1080) has been shown to participate in the generation of angiostatin from plasmin. In this case, the reduction of the plasmin disulfide bonds triggers the proteolytic cleavage of plasmin which yields the angiogenesis inhibitor, angiostatin (Stathakis, P. et al. (1997) J. Biol. Chem. 272:20641-20645). Low levels of reduced sulfhydryl groups in plasma has been associated with rheumatoid arthritis. The failure of these sulfhydryl groups to scavenge active oxygen species (e.g., hydrogen peroxide produced by activated neutrophils) results in oxidative damage to surrounding tissues and the resulting inflammation (Hall, N.D. et al. (1994) Rheumatol. Int. 4:35-38).

Another example of the importance of redox reactions in cell metabolism is the degradation of saturated and unsaturated fatty acids by mitochondrial and peroxisomal beta-oxidation enzymes

which sequentially remove two-carbon units from Coenzyme A (CoA)-activated fatty acids. The main beta-oxidation pathway degrades both saturated and unsaturated fatty acids while the auxiliary pathway performs additional steps required for the degradation of unsaturated fatty acids.

The pathways of mitchondrial and peroxisomal beta-oxidation use similar enzymes, but have different substrate specificities and functions. Mitochondria oxidize short-, medium-, and long-chain fatty acids to produce energy for cells. Mitochondrial beta-oxidation is a major energy source for cardiac and skeletal muscle. In liver, it provides ketone bodies to the peripheral circulation when glucose levels are low as in starvation, endurance exercise, and diabetes (Eaton, S. et al. (1996) Biochem. J. 320:345-357). Peroxisomes oxidize medium-, long-, and very-long-chain fatty acids, dicarboxylic fatty acids, branched fatty acids, prostaglandins, xenobiotics, and bile acid intermediates. The chief roles of peroxisomal beta-oxidation are to shorten toxic lipophilic carboxylic acids to facilitate their excretion and to shorten very-long-chain fatty acids prior to mitochondrial beta-oxidation (Mannaerts, G.P. and P.P. Van Veldhoven (1993) Biochimie 75:147-158).

The auxiliary beta-oxidation enzyme 2,4-dienoyl-CoA reductase catalyzes the following reaction:

trans-2, cis/trans-4-dienoyl-CoA + NADPH + $H^+$ ---> trans-3-enoyl-CoA + $NADP^+$

This reaction removes even-numbered double bonds from unsaturated fatty acids prior to their entry into the main beta-oxidation pathway (Koivuranta, K.T. et al. (1994) Biochem. J. 304:787-792). The enzyme may also remove odd-numbered double bonds from unsaturated fatty acids (Smeland, T.E. et al. (1992) Proc. Natl. Acad. Sci. USA 89:6673-6677).

Rat 2,4-dienoyl-CoA reductase is located in both mitochondria and peroxisomes (Dommes, V. et al. (1981) J. Biol. Chem. 256:8259-8262). Two immunologically different forms of rat mitochondrial enzyme exist with molecular masses of 60 kDa and 120 kDa (Hakkola, E.H. and J.K. Hiltunen (1993) Eur. J. Biochem. 215:199-204). The 120 kDa mitochondrial rat enzyme is synthesized as a 335 amino acid precursor with a 29 amino acid N-terminal leader peptide which is cleaved to form the mature enzyme (Hirose, A. et al. (1990) Biochim. Biophys. Acta 1049:346-349). A human mitochondrial enzyme 83% similar to rat enzyme is synthesized as a 335 amino acid residue precursor with a 19 amino acid N-terminal leader peptide (Koivuranta, supra). These cloned human and rat mitochondrial enzymes function as homotetramers (Koivuranta, supra). A Saccharomyces cerevisiae peroxisomal 2,4-dienoyl-CoA reductase is 295 amino acids long, contains a C-terminal peroxisomal targeting signal, and functions as a homodimer (Coe, J.G.S. et al. (1994) Mol. Gen. Genet. 244:661-672; and Gurvitz, A. et al. (1997) J. Biol. Chem. 272:22140-22147). All 2,4-dienoyl-CoA reductases have a fairly well conserved NADPH binding site motif (Koivuranta,

supra).

The main pathway beta-oxidation enzyme enoyl-CoA hydratase catalyzes the reaction:

2-trans-enoyl-CoA + $H_2O$ <---> 3-hydroxyacyl-CoA

This reaction hydrates the double bond between C-2 and C-3 of 2-trans-enoyl-CoA, which is generated from saturated and unsaturated fatty acids (Engel, C.K. et al. (1996) EMBO J. 15:5135-5145). This step is downstream from the step catalyzed by 2,4-dienoyl-reductase. Different enoyl-CoA hydratases act on short-, medium-, and long-chain fatty acids (Eaton, supra). Mitochondrial and peroxisomal enoyl-CoA hydratases occur as both mono-functional enzymes and as part of multi-functional enzyme complexes. Human liver mitochondrial short-chain enoyl-CoA hydratase is synthesized as a 290 amino acid precursor with a 29 amino acid N-terminal leader peptide (Kanazawa, M. et al. (1993) Enzyme Protein 47:9-13; and Janssen, U. et al. (1997) Genomics 40:470-475). Rat short-chain enoyl-CoA hydratase is 87% identical to the human sequence in the mature region of the protein and functions as a homohexamer (Kanazawa, supra; and Engel, supra). A mitochondrial trifunctional protein exists that has long-chain enoyl-CoA hydratase, 3-hydroxyacyl-CoA dehydrogenase, and long-chain 3-oxothiolase activities (Eaton, supra). In human peroxisomes, enoyl-CoA hydratase activity is found in both a 327 amino acid residue mono-functional enzyme and as part of a multi-functional enzyme, also known as bifunctional enzyme, which possesses enoyl-CoA hydratase, enoyl-CoA isomerase, and 3-hydroxyacyl-CoA hydrogenase activities (FitzPatrick, D.R. et al. (1995) Genomics 27:457-466; and Hoefler, G. et al. (1994) Genomics 19:60-67). A 339 amino acid residue human protein with short-chain enoyl-CoA hydratase activity also acts as an AU-specific RNA binding protein (Nakagawa, J. et al. (1995) Proc. Natl. Acad. Sci. USA 92:2051-2055). All enoyl-CoA hydratases share homology near two active site glutamic acid residues, with 17 amino acid residues that are highly conserved (Wu, W.-J. et al. (1997) Biochemistry 36:2211-2220).

Inherited deficiencies in mitochondrial and peroxisomal beta-oxidation enzymes are associated with severe diseases, some of which manifest soon after birth and lead to death within a few years. Mitochondrial beta-oxidation associated deficiencies include, e.g., carnitine palmitoyl transferase and carnitine deficiency, very-long-chain acyl-CoA dehydrogenase deficiency, medium-chain acyl-CoA dehydrogenase deficiency, short-chain acyl-CoA dehydrogenase deficiency, electron transport flavoprotein and electron transport flavoprotein:ubiquinone oxidoreductase deficiency, trifunctional protein deficiency, and short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency (Eaton, supra). Mitochondrial trifunctional protein (including enoyl-CoA hydratase) deficient patients have reduced long-chain enoyl-CoA hydratase activities and suffer from non-ketotic hypoglycemia, sudden infant death syndrome, cardiomyopathy, hepatic dysfunction, and muscle

weakness, and may die at an early age (Eaton, supra).

Defects in mitochondrial beta-oxidation are associated with Reye's syndrome, a disease characterized by hepatic dysfunction and encephalopathy that sometimes follows viral infection in children. Reye's syndrome patients may have elevated serum levels of free fatty acids (Cotran, R.S. et al. (1994) Robbins Pathologic Basis of Disease, W.B. Saunders Co., Philadelphia PA, p.866). Patients with mitochondrial short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency and medium-chain 3-hydroxyacyl-CoA dehydrogenase deficiency also exhibit Reye-like illnesses (Eaton, supra; and Egidio, R.J. et al. (1989) Am. Fam. Physician 39:221-226).

Inherited conditions associated with peroxisomal beta-oxidation include Zellweger syndrome, neonatal adrenoleukodystrophy, infantile Refsum's disease, acyl-CoA oxidase deficiency, peroxisomal thiolase deficiency, and bifunctional protein deficiency (Suzuki, Y. et al. (1994) Am. J. Hum. Genet. 54:36-43; Hoefler, supra). Patients with peroxisomal bifunctional enzyme deficiency, including that of enoyl-CoA hydratase, suffer from hypotonia, seizures, psychomotor defects, and defective neuronal migration; accumulate very-long-chain fatty acids; and typically die within a few years of birth (Watkins, P.A. et al. (1989) J. Clin. Invest. 83:771-1277).

Peroxisomal beta-oxidation is impaired in cancerous tissue. Although neoplastic human breast epithelial cells have the same number of peroxisomes as do normal cells, fatty acyl-CoA oxidase activity is lower than in control tissue (el Bouhtoury, F. et al. (1992) J. Pathol. 166:27-35). Human colon carcinomas have fewer peroxisomes than normal colon tissue and have lower fatty-acyl-CoA oxidase and bifunctional enzyme (including enoyl-CoA hydratase) activities than normal tissue (Cable, S. et al. (1992) Virchows Arch. B Cell Pathol. Incl. Mol. Pathol. 62:221-226).

6-phosphogluconate dehydrogenase (6-PGDH) catalyses the $NADP^+$-dependent oxidative decarboxylation of 6-phosphogluconate to ribulose 5-phosphate with the production of NADPH. The absence or inhibition of 6-PGDH results in the accumulation of 6-phosphogluconate to toxic levels in eukaryotic cells. 6-PGDH is the third enzyme of the pentose phosphate pathway (PPP) and is ubiquitous in nature. In some heterofermentatative species, NAD+ is used as a cofactor with the subsequent production of NADH.

The reaction proceeds through a 3-keto intermediate which is decarboxylated to give the enol of ribulose 5-phosphate, then converted to the keto product following tautomerization of the enol (Berdis A.J. and P.F. Cook (1993) Biochemistry 32:2041-2046). 6-PGDH activity is regulated by the inhibitory effect of NADPH, and the activating effect of 6-phosphogluconate (Rippa, M. et al. (1998) Biochim. Biophys. Acta 1429:83-92). Deficiencies in 6-PGDH activity have been linked to chronic hemolytic anemia.

The targeting of specific forms of 6-PGDH (e.g., enzymes found in trypanosomes) has been suggested as a means for controlling parasitic infections (Tetaud, E. et al. (1999) Biochem. J.

338:55-60). For example, the T. brucei enzyme is markedly more sensitive to inhibition by the substrate analogue 6-phospho-2-deoxygluconate and the coenzyme analogue adenosine 2',5'-bisphosphate, compared to the mammalian enzyme (Hanau, S. et al. (1996) Eur. J. Biochem. 240:592-599).

Ribonucleotide diphosphate reductase catalyzes the reduction of ribonucleotide diphosphates (i.e., ADP, GDP, CDP, and UDP) to their corresponding deoxyribonucleotide diphosphates (i.e., dADP, dGDP, dCDP, and dUDP) which are used for the synthesis of DNA. Ribonucleotide diphosphate reductase thereby performs a crucial role in the de novo synthesis of deoxynucleotide precursors. Deoxynucleotides are also produced from deoxynucleosides by nucleoside kinases via the salvage pathway.

Mammalian ribonucleotide diphosphate reductase comprises two components, an effector-binding component (E) and a non-heme iron component (F). Component E binds the nucleoside triphosphate effectors while component F contains the iron radical necessary for catalysis. Molecular weight determinations of the E and F components, as well as the holoenzyme, vary according to the methods used in purification of the proteins and the particular laboratory. Component E is approximately 90-100 kDa, component F is approximately 100-120 kDa, and the holoenzyme is 200-250 kDa.

Ribonucleotide diphosphate reductase activity is adversely effected by iron chelators, such as thiosemicarbazones, as well as EDTA. Deoxyribonucleotide diphosphates also appear to be negative allosteric effectors of ribonucleotide diphosphate reductase. Nucleotide triphosphates (both ribo- and deoxyribo-) appear to stimulate the activity of the enzyme. 3-methyl-4-nitrophenol, a metabolite of widely used organophosphate pesticides, is a potent inhibitor of ribonucleotide diphosphate reductase in mammalian cells. Some evidence suggests that ribonucleotide diphosphate reductase activity in DNA virus (e.g., herpes virus) -infected cells and in cancer cells is less sensitive to regulation by allosteric regulators and a correlation exists between high ribonucleotide diphosphate reductase activity levels and high rates of cell proliferation (e.g., in hepatomas). This observation suggests that virus-encoded ribonucleotide diphosphate reductases, and those present in cancer cells, are capable of maintaining an increased supply deoxyribonucleotide pool for the production of virus genomes or for the increased DNA synthesis which characterizes cancers cells. Ribonucleotide diphosphate reductase is thus a target for therapeutic intervention (Nutter, L.M. and Y.-C. Cheng (1984) Pharmac. Ther. 26:191-207; and Wright, J.A. (1983) Pharmac. Ther. 22:81-102).

Dihydrodiol dehydrogenases (DD) are monomeric, $NAD(P)^+$-dependent, 34-37 kDa enzymes responsible for the detoxification of *trans*-dihydrodiol and *anti*-diol epoxide metabolites of polycyclic aromatic hydrocarbons (PAH) such as benzo[*a*]yrene, benz[*a*]anthracene, 7-methyl-benz[*a*]anthracene, 7,12-dimethyl-benz[*a*]anthracene, chrysene, and 5-methyl-chrysene. In

mammalian cells, an environmental PAH toxin such as benzo[*a*]yrene is initially epoxidated by a microsomal cytochrome P450 to yield 7*R*,8*R*-arene-oxide and subsequently (-)-7*R*,8*R*-dihydrodiol ((-)-*trans*-7,8-dihydroxy-7,8-dihydrobenzo[*a*]pyrene or (-)-*trans*-B[*a*]P-diol) This latter compound is further transformed to the *anti*-diol epoxide of benzo[*a*]pyrene (i.e., (±)-anti-7$\beta$,8$\alpha$-dihydroxy-9$\alpha$,10$\alpha$-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene), by the same enzyme or a different enzyme, depending on the species. This resulting *anti*-diol epoxide of benzo[*a*]yrene, or the corresponding derivative from another PAH compound, is highly mutagenic.

DD efficiently oxidizes the precursor of the *anti*-diol epoxide (i.e., *trans*-dihydrodiol) to transient catechols which auto-oxidize to quinones, also producing hydrogen peroxide and semiquinone radicals. This reaction prevents the formation of the highly carcinogenic *anti*-diol. *Anti*-diols are not themselves substrates for DD yet the addition of DD to a sample comprising an *anti*-diol compound results in a significant decrease in the induced mutation rate observed in the Ames test. In this instance, DD is able to bind to and sequester the *anti*-diol, even though it is not oxidized. Whether through oxidation or sequestration, DD plays an important role in the detoxification of metabolites of xenobiotic polycyclic compounds (Penning, T.M. (1993) Chemico-Biological Interactions 89:1-34).

15-oxoprostaglandin 13-reductase (PGR) and 15-hydroxyprostaglandin dehydrogenase (15-PGDH) are enzymes present in the lung that are responsible for degrading circulating prostaglandins. Oxidative catabolism via passage through the pulmonary system is a common means of reducing the concentration of circulating prostaglandins. 15-PGDH oxidizes the 15-hydroxyl group of a variety of prostaglandins to produce the corresponding 15-oxo compounds. The 15-oxo derivatives usually have reduced biological activity compared to the 15-hydroxyl molecule. PGR further reduces the 13,14 double bond of the 15-oxo compound which typically leads to a further decrease in biological activity. PGR is a monomer with a molecular weight of approximately 36 kDa. The enzyme requires NADH or NADPH as a cofactor with a preference for NADH. The 15-oxo derivatives of prostaglandins $PGE_1$, $PGE_2$, and $PGE_{2\alpha}$ are all substrates for PGR; however, the non-derivatized prostaglandins (i.e., $PGE_1$, $PGE_2$, and $PGE_{2\alpha}$) are not substrates (Ensor, C.M. et al. (1998) Biochem. J. 330:103-108).

15-PGDH and PGR also catalyze the metabolism of lipoxin $A_4$ ($LXA_4$). Lipoxins (LX) are autacoids, lipids produced at the sites of localized inflammation, which down-regulate polymorphonuclear leukocyte (PMN) function and promote resolution of localized trauma. Lipoxin production is stimulated by the administration of aspirin in that cells displaying cyclooxygenase II (COX II) that has been acetylated by aspirin and cells that possess 5-lipoxygenase (5-LO) interact and produce lipoxin. 15-PGDH generates 15-oxo-$LXA_4$ with PGR further converting the 15-oxo compound to 13,14-dihydro-15-oxo-$LXA_4$ (Clish, C.B. et al. (2000) J. Biol. Chem.

275:25372-25380). This finding suggests a broad substrate specificity of the prostaglandin dehydrogenases and has implications for these enzymes in drug metabolism and as targets for therapeutic intervention to regulate inflammation.

The GMC (glucose-methanol-choline) oxidoreductase family of enzymes was defined based on sequence alignments of Drosophila melanogaster glucose dehydrogenase, Escherichia coli choline dehydrogenase, Aspergillus niger glucose oxidase, and Hansenula polymorpha methanol oxidase. Despite their different sources and substrate specificities, these four flavoproteins are homologous, being characterized by the presence of several distinctive sequence and structural features. Each molecule contains a canonical ADP-binding, beta-alpha-beta mononucleotide-binding motif close to the amino terminus. This fold comprises a four-stranded parallel beta-sheet sandwiched between a three-stranded antiparallel beta-sheet and alpha-helices. Nucleotides bind in similar positions relative to this chain fold (Cavener, D.R. (1992) J. Mol. Biol. 223:811-814; Wierenga, R.K. et al. (1986) J. Mol. Biol. 187:101-107). Members of the GMC oxidoreductase family also share a consensus sequence near the central region of the polypeptide. Additional members of the GMC oxidoreductase family include cholesterol oxidases from Brevibacterium sterolicum and Streptomyces; and an alcohol dehydrogenase from Pseudomonas oleovorans (Cavener, D.R., supra; Henikoff, S. and J.G. Henikoff (1994) Genomics 19:97-107; van Beilen, J.B. et al. (1992) Mol. Microbiol. 6:3121-3136).

IMP dehydrogenase and GMP reductase are two oxidoreductases which share many regions of sequence similarity. IMP dehydrogenase (EC 1.1.1.205) catalyes the NAD-dependent reduction of IMP (inosine monophosphate) into XMP (xanthine monophosphate) as part of de novo GTP biosynthesis (Collart, F.R. and E. Huberman (1988) J. Biol. Chem. 263:15769-15772). GMP reductase catalyzes the NADPH-dependent reductive deamination of GMP into IMP, helping to maintain the intracellular balance of adenine and guanine nucleotides (Andrews, S.C. and J.R. Guest (1988) Biochem. J. 255:35-43).

Pyridine nucleotide-disulphide oxidoreductases are FAD flavoproteins involved in the transfer of reducing equivalents from FAD to a substrate. These flavoproteins contain a pair of redox-active cysteines contained within a consensus sequence which is characteristic of this protein family (Kurlyan, J. et al. (1991) Nature 352:172-174). Members of this family of oxidoreductases include glutathione reductase (EC 1.6.4.2); thioredoxin reductase of higher eukaryotes (EC 1.6.4.5); trypanothione reductase (EC 1.6.4.8); lipoamide dehydrogenase (EC 1.8.1.4), the E3 component of alpha-ketoacid dehydrogenase complexes; and mercuric reductase (EC 1.16.1.1).

Transferases

Transferases are enzymes that catalyze the transfer of molecular groups. The reaction may involve an oxidation, reduction, or cleavage of covalent bonds, and is often specific to a substrate or to particular sites on a type of substrate. Transferases participate in reactions essential to such

functions as synthesis and degradation of cell components, and regulation of cell functions including cell signaling, cell proliferation, inflammation, apoptosis, secretion and excretion. Transferases are involved in key steps in disease processes involving these functions. Transferases are frequently classified according to the type of group transferred. For example, methyl transferases transfer one-carbon methyl groups, amino transferases transfer nitrogenous amino groups, and similarly denominated enzymes transfer aldehyde or ketone, acyl, glycosyl, alkyl or aryl, isoprenyl, saccharyl, phosphorous-containing, sulfur-containing, or selenium-containing groups, as well as small enzymatic groups such as Coenzyme A.

Acyl transferases include peroxisomal carnitine octanoyl transferase, which is involved in the fatty acid beta-oxidation pathway, and mitochondrial carnitine palmitoyl transferases, involved in fatty acid metabolism and transport. Choline O-acetyl transferase catalyzes the biosynthesis of the neurotransmitter acetylcholine. N-acyltransferase enzymes catalyze the transfer of an amino acid conjugate to an activated carboxylic group. Endogenous compounds and xenobiotics are activated by acyl-CoA synthetases in the cytosol, microsomes, and mitochondria. The acyl-CoA intermediates are then conjugated with an amino acid (typically glycine, glutamine, or taurine, but also ornithine, arginine, histidine, serine, aspartic acid, and several dipeptides) by N-acyltransferases in the cytosol or mitochondria to form a metabolite with an amide bond. One well-characterized enzyme of this class is the bile acid-CoA:amino acid N-acyltransferase (BAT) responsible for generating the bile acid conjugates which serve as detergents in the gastrointestinal tract (Falany, C. N. et al. (1994) J. Biol. Chem. 269:19375-9; Johnson, M. R. et al. (1991) J. Biol. Chem. 266:10227-33). BAT is also useful as a predictive indicator for prognosis of hepatocellular carcinoma patients after partial hepatectomy (Furutani, M. et al. (1996) Hepatology 24:1441-5).

Acetyltransferases

Acetyltransferases have been extensively studied for their role in histone acetylation. Histone acetylation results in the relaxing of the chromatin structure in eukaryotic cells, allowing transcription factors to gain access to promoter elements of the DNA templates in the affected region of the genome (or the genome in general). In contrast, histone deacetylation results in a reduction in transcription by closing the chromatin structure and limiting access of transcription factors. To this end, a common means of stimulating cell transcription is the use of chemical agents that inhibit the deacetylation of histones (e.g., sodium butyrate), resulting in a global (albeit artifactual) increase in gene expression. The modulation of gene expression by acetylation also results from the acetylation of other proteins, including but not limited to, p53, GATA-1, MyoD, ACTR, TFIIE, TFIIF and the high mobility group proteins (HMG). In the case of p53, acetylation results in increased DNA binding, leading to the stimulation of transcription of genes regulated by p53. The prototypic histone

acetylase (HAT) is Gcn5 from Saccharomyces cerevisiae. Gcn5 is a member of a family of acetylases that includes Tetrahymena p55, human Gcn5, and human p300/CBP. Histone acetylation is reviewed in (Cheung, W.L. et al. (2000) Curr. Opin. Cell Biol. 12:326-333 and Berger, S.L (1999) Curr. Opin. Cell Biol. 11:336-341). Some acetyltransferase enzymes possess the alpha/beta hydrolase fold (Center of Applied Molecular Engineering Inst. of Chemistry and Biochemistry - University of Salzburg, http://predict.sanger.ac.uk/irbm-course97/Docs/ms/) common to several other major classes of enzymes, including but not limited to, acetylcholinesterases and carboxylesterases (Structural Classification of Proteins, http://scop.mrc-lmb.cam.ac.uk/scop/index.html).

N-acetyltransferases are cytosolic enzymes which utilize the cofactor acetyl-coenzyme A (acetyl-CoA) to transfer the acetyl group to aromatic amines and hydrazine containing compounds. In humans, there are two highly similar N-acetyltransferase enzymes, NAT1 and NAT2; mice appear to have a third form of the enzyme, NAT3. The human forms of N-acetyltransferase have independent regulation (NAT1 is widely-expressed, whereas NAT2 is in liver and gut only) and overlapping substrate preferences. Both enzymes appear to accept most substrates to some extent, but NAT1 does prefer some substrates (para-aminobenzoic acid, para-aminosalicylic acid, sulfamethoxazole, and sulfanilamide), while NAT2 prefers others (isoniazid, hydralazine, procainamide, dapsone, aminoglutethimide, and sulfamethazine). A recently isolated human gene, tubedown-1, is homologous to the yeast NAT-1 N-acetyltransferases and encodes a protein associated with acetyltransferase activity. The expression patterns of tubedown-1 suggest that it may be involved in regulating vascular and hematopoietic development (Gendron, R.L. et al. (2000) Dev. Dyn. 218:300-315).

Amino transferases comprise a family of pyridoxal 5'-phosphate (PLP) -dependent enzymes that catalyze transformations of amino acids. Amino transferases play key roles in protein synthesis and degradation, and they contribute to other processes as well. For example, GABA aminotransferase (GABA-T) catalyzes the degradation of GABA, the major inhibitory amino acid neurotransmitter. The activity of GABA-T is correlated to neuropsychiatric disorders such as alcoholism, epilepsy, and Alzheimer's disease (Sherif, F.M. and Ahmed, S.S. (1995) Clin. Biochem. 28:145-154). Other members of the family include pyruvate aminotransferase, branched-chain amino acid aminotransferase, tyrosine aminotransferase, aromatic aminotransferase, alanine:glyoxylate aminotransferase (AGT), and kynurenine aminotransferase (Vacca, R.A. et al. (1997) J. Biol. Chem. 272:21932-21937). Kynurenine aminotransferase catalyzes the irreversible transamination of the L-tryptophan metabolite L-kynurenine to form kynurenic acid. The enzyme may also catalyzes the reversible transamination reaction between L-2-aminoadipate and 2-oxoglutarate to produce 2-oxoadipate and L-glutamate. Kynurenic acid is a putative modulator of glutamatergic

neurotransmission, thus a deficiency in kynurenine aminotransferase may be associated with pleiotropic effects (Buchli, R. et al. (1995) J. Biol. Chem. 270:29330-29335).

Glycosyl transferases include the mammalian UDP-glucouronosyl transferases, a family of membrane-bound microsomal enzymes catalyzing the transfer of glucouronic acid to lipophilic substrates in reactions that play important roles in detoxification and excretion of drugs, carcinogens, and other foreign substances. Another mammalian glycosyl transferase, mammalian UDP-galactose-ceramide galactosyl transferase, catalyzes the transfer of galactose to ceramide in the synthesis of galactocerebrosides in myelin membranes of the nervous system. The UDP-glycosyl transferases share a conserved signature domain of about 50 amino acid residues (PROSITE: PDOC00359, http://expasy.hcuge.ch/sprot/prosite.html).

Methyl transferases are involved in a variety of pharmacologically important processes. Nicotinamide N-methyl transferase catalyzes the N-methylation of nicotinamides and other pyridines, an important step in the cellular handling of drugs and other foreign compounds. Phenylethanolamine N-methyl transferase catalyzes the conversion of noradrenalin to adrenalin. 6-O-methylguanine-DNA methyl transferase reverses DNA methylation, an important step in carcinogenesis. Uroporphyrin-III C-methyl transferase, which catalyzes the transfer of two methyl groups from S-adenosyl-L-methionine to uroporphyrinogen III, is the first specific enzyme in the biosynthesis of cobalamin, a dietary enzyme whose uptake is deficient in pernicious anemia. Protein-arginine methyl transferases catalyze the posttranslational methylation of arginine residues in proteins, resulting in the mono- and dimethylation of arginine on the guanidino group. Substrates include histones, myelin basic protein, and heterogeneous nuclear ribonucleoproteins involved in mRNA processing, splicing, and transport. Protein-arginine methyl transferase interacts with proteins upregulated by mitogens, with proteins involved in chronic lymphocytic leukemia, and with interferon, suggesting an important role for methylation in cytokine receptor signaling (Lin, W.-J. et al. (1996) J. Biol. Chem. 271:15034-15044; Abramovich, C. et al. (1997) EMBO J. 16:260-266; and Scott, H. S. et al. (1998) Genomics 48:330-340).

Phospho transferases catalyze the transfer of high-energy phosphate groups and are important in energy-requiring and -releasing reactions. The metabolic enzyme creatine kinase catalyzes the reversible phosphate transfer between creatine/creatine phosphate and ATP/ADP. Glycocyamine kinase catalyzes phosphate transfer from ATP to guanidoacetate, and arginine kinase catalyzes phosphate transfer from ATP to arginine. A cysteine-containing active site is conserved in this family (PROSITE: PDOC00103).

Prenyl transferases are heterodimers, consisting of an alpha and a beta subunit, that catalyze the transfer of an isoprenyl group. The Ras farnesyltransferase (FTase) enzyme transfers a farnesyl moiety from cytosolic farnesylpyrophosphate to a cysteine residue at the carboxyl terminus of the

Ras oncogene protein. This modification is required to anchor Ras to the cell membrane so that it can perform its role in signal transduction. FTase inhibitors block Ras function and demonstrate antitumor activity (Buolamwini, J.K. (1999) Curr. Opin. Chem. Biol. 3:500-509). Ftase, which shares structural similarity with geranylgeranyl transferase, or Rab GG transferase, prenylates Rab proteins, allowing them to perform their roles in regulating vesicle transport (Seabra, M.C. (1996) J. Biol. Chem. 271:143913-24404).

Saccharyl transferases are glycating enzymes involved in a variety of metabolic processes. Oligosaccharyl transferase-48, for example, is a receptor for advanced glycation endproducts, which accumulate in vascular complications of diabetes, macrovascular disease, renal insufficiency, and Alzheimer's disease (Thornalley, P. J. (1998) Cell Mol. Biol. (Noisy-Le-Grand) 44:1013-1023).

Coenzyme A (CoA) transferase catalyzes the transfer of CoA between two carboxylic acids. Succinyl CoA:3-oxoacid CoA transferase, for example, transfers CoA from succinyl-CoA to a recipient such as acetoacetate. Acetoacetate is essential to the metabolism of ketone bodies, which accumulate in tissues affected by metabolic disorders such as diabetes (PROSITE: PDOC00980).

Transglutaminase transferases (Tgases) are $Ca^{2+}$ dependent enzymes capable of forming isopeptide bonds by catalyzing the transfer of the $\gamma$-carboxy group from protein-bound glutamine to the $\epsilon$-amino group of protein-bound lysine residues or other primary amines. Tgases are the enzymes responsible for the cross-linking of cornified envelope (CE), the highly insoluble protein structure on the surface of corneocytes, into a chemically and mechanically resistant protein polymer. Seven known human Tgases have been identified. Individual transglutaminase gene products are specialized in the cross-linking of specific proteins or tissue structures, such as factor XIIIa which stabilizes the fibrin clot in hemostasis, prostrate transglutaminase which functions in semen coagulation, and tissue transglutaminase which is involved in GTP-binding in receptor signaling. Four (Tgases 1, 2, 3, and X) are expressed in terminally differentiating epithelia such as the epidermis. Tgases are critical for the proper cross-linking of the CE as seen in the pathology of patients suffering from one form of the skin diseases referred to as congenital ichthyosis which has been linked to mutations in the keratinocyte transglutaminase ($TG_K$) gene (Nemes, Z. et al., (1999) Proc. Natl. Acad. Sci. U.S.A. 96:8402-8407, Aeschlimann, D. et al., (1998) J. Biol. Chem. 273:3452-3460.)

Hydrolases

Hydrolases are a class of enzymes that catalyze the cleavage of various covalent bonds in a substrate by the introduction of a molecule of water. The reaction involves a nucleophilic attack by the water molecule's oxygen atom on a target bond in the substrate. The water molecule is split across the target bond, breaking the bond and generating two product molecules. Hydrolases

participate in reactions essential to such functions as synthesis and degradation of cell components, and for regulation of cell functions including cell signaling, cell proliferation, inflamation, apoptosis, secretion and excretion. Hydrolases are involved in key steps in disease processes involving these functions. Hydrolytic enzymes, or hydrolases, may be grouped by substrate specificity into classes including phosphatases, peptidases, lysophospholipases, phosphodiesterases, glycosidases, glyoxalases, aminohydrolases, carboxylesterases, sulfatases, phosphohydrolases, nucleotidases, lysozymes, and many others.

Phosphatases hydrolytically remove phosphate groups from proteins, an energy-providing step that regulates many cellular processes, including intracellular signaling pathways that in turn control cell growth and differentiation, cell-cell contact, the cell cycle, and oncogenesis.

Peptidases, also called proteases, cleave peptide bonds that form the backbone of peptide or protein chains. Proteolytic processing is essential to cell growth, differentiation, remodeling, and homeostasis as well as inflammation and the immune response. Since typical protein half-lives range from hours to a few days, peptidases are continually cleaving precursor proteins to their active form, removing signal sequences from targeted proteins, and degrading aged or defective proteins. Peptidases function in bacterial, parasitic, and viral invasion and replication within a host. Examples of peptidases include trypsin and chymotrypsin (components of the complement cascade and the blood-clotting cascade) lysosomal cathepsins, calpains, pepsin, renin, and chymosin (Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York, NY, pp. 1-5).

Lysophospholipases (LPLs) regulate intracellular lipids by catalyzing the hydrolysis of ester bonds to remove an acyl group, a key step in lipid degradation. Small LPL isoforms, approximately 15-30 kD, function as hydrolases; larger isoforms function both as hydrolases and transacylases. A particular substrate for LPLs, lysophosphatidylcholine, causes lysis of cell membranes. LPL activity is regulated by signaling molecules important in numerous pathways, including the inflammatory response.

The phosphodiesterases catalyze the hydrolysis of one of the two ester bonds in a phosphodiester compound. Phosphodiesterases are therefore crucial to a variety of cellular processes. Phosphodiesterases include DNA and RNA endo- and exo-nucleases, which are essential to cell growth and replication as well as protein synthesis. Endonuclease V (deoxyinosine 3′-endonuclease) is an example of a type II site-specific deoxyribonuclease, a putative DNA repair enzyme that cleaves DNAs containing hypoxanthine, uracil, or mismatched bases. Escherichia coli endonuclease V has been shown to cleave DNA containing deoxyxanthosine at the second phosphodiester bond 3′ to deoxyxanthosine, generating a 3′-hydroxyl and a 5′-phosphoryl group at the nick site (He, B. et al. (2000) Mutat. Res. 459:109-114). It has been suggested that Escherichia coli endonuclease V plays a

role in the removal of deaminated guanine, i.e., xanthine, from DNA, thus helping to protect the cell against the mutagenic effects of nitrosative deamination (Schouten, K.A. and Weiss, B. (1999) Mutat. Res. 435:245-254). In eukaryotes, the process of tRNA splicing requires the removal of small tRNA introns that interrupt the anticodon loop 1 base 3′ to the anticodon. This process requires the stepwise action of an endonuclease, a ligase, and a phosphotransferase (Hong, L. et al. (1998) Science 280:279-284). Ribonuclease P (RNase P) is a ubiquitous RNA processing endonuclease that is required for generating the mature tRNA 5′-end during the tRNA splicing process. This is accomplished through the catalysis of the cleavage of P-3′O bonds to produce 5′-phosphate and 3′-hydroxyl end groups at a specific site on pre-tRNA. Catalysis by RNase P is absolutely dependent on divalent cations such as $Mg^{2+}$ or $Mn^{2+}$ (Kurz, J.C. et al. (2000) Curr. Opin. Chem. Biol. 4:553-558). Substrate recognition mechanisms of RNase P are well conserved among eukaryotes and bacteria (FENZMi, S. et al. (1998) Science 280:284-286). In S. cerevisiae, POP1 ('processing of precursor RNAs') encodes a protein component of both RNase P and RNase MRP, another RNA processing protein. Mutations in yeast POP1 are lethal (Lygerou, Z. et al. (1994) Genes Dev. 8:1423-1433). Another phosphodiesterase, acid sphingomyelinase, hydrolyzes the membrane phospholipid sphingomyelin to ceramide and phosphorylcholine. Phosphorylcholine functions in synthesis of phosphatidylcholine, which is involved in intracellular signaling pathways. Ceramide is an essential precursor for the generation of gangliosides, membrane lipids found in high concentration in neural tissue. Defective acid sphingomyelinase phosphodiesterase leads to Niemann-Pick disease.

Glycosidases catalyze the cleavage of hemiacetyl bonds of glycosides, which are compounds that contain one or more sugar. Mammalian lactase-phlorizin hydrolase, for example, is an intestinal enzyme that splits lactose. Mammalian beta-galactosidase removes the terminal galactose from gangliosides, glycoproteins, and glycosaminoglycans, and deficiency of this enzyme is associated with a gangliosidosis known as Morquio disease type B (PROSITE PCDOC00910). Vertebrate lysosomal alpha-glucosidase, which hydrolyzes glycogen, maltose, and isomaltose, and vertebrate intestinal sucrase-isomaltase, which hydrolyzes sucrose, maltose, and isomaltose, are widely distributed members of this family with highly conserved sequences at their active sites.

The glyoxylase system is involved in gluconeogenesis, the production of glucose from storage compounds in the body. It consists of glyoxylase I, which catalyzes the formation of S-D-lactoylglutathione from methyglyoxal, a side product of triose-phosphate energy metabolism, and glyoxylase II, which hydrolyzes S-D-lactoylglutathione to D-lactic acid and reduced glutathione. Glyoxylases are involved in hyperglycemia, non-insulin-dependent diabetes mellitus, the detoxification of bacterial toxins, and in the control of cell proliferation and microtubule assembly.

NG,NG-dimethylarginine dimethylaminohydrolase (DDAH) is an enzyme that hydrolyzes the endogenous nitric oxide synthase (NOS) inhibitors, NG-monomethyl-arginine and NG,NG-dimethyl-

L-arginine, to L-citrulline. Inhibiting DDAH can cause increased intracellular concentration of NOS inhibitors to levels sufficient to inhibit NOS. Therefore, DDAH inhibition may provide a method of NOS inhibition, and changes in the activity of DDAH could play a role in pathophysiological alterations in nitric oxide generation (MacAllister, R.J. et al. (1996) Br. J. Pharmacol. 119:1533-1540). DDAH was found in neurons displaying cytoskeletal abnormalities and oxidative stress in Alzheimer's disease. In age-matched control cases, DDAH was not found in neurons. This suggests that oxidative stress- and nitric oxide-mediated events play a role in the pathogenesis of Alzheimer's disease (Smith, M.A. et al. (1998) Free Rad. Biol. Med. 25:898-902).

Acyl-CoA thioesterase is another member of the carboxylesterase family (Alexson, S.E. et al. (1993) Eur. J. Biochem. 214:719-727). Evidence suggests that acyl-CoA thioesterase has a regulatory role in steroidogenic tissues (Finkielstein, C. et al. (1998) Eur. J. Biochem. 256:60-66).

The alpha/beta hydrolase protein fold is common to several hydrolases of diverse phylogenetic origin and catalytic function. Enzymes with the alpha/beta hydrolase fold have a common core structure consisting of eight beta-sheets connected by alpha-helices. The most conserved structural feature of this fold is the loops of the nucleophile-histidine-acid catalytic triad. The histidine in the catalytic triad is completely conserved, while the nucleophile and acid loops accommodate more than one type of amino acid (Ollis, D.L. et al. (1992) Protein Eng. 5:197-211).

Sulfatases are members of a highly conserved gene family that share extensive sequence homology and a high degree of structural similarity. Sulfatases catalyze the cleavage of sulfate esters. To perform this function, sulfatases undergo a unique post-translational modification in the endoplasmic reticulum that involves the oxidation of a conserved cysteine residue. A human disorder called multiple sulfatase deficiency is due to a defect in this post-translational modification step, leading to inactive sulfatases (Recksiek, M. et al. (1998) J. Biol. Chem. 273:6096-6103).

Phosphohydrolases are enzymes that hydrolyze phosphate esters. Some phosphohydrolases contain a mutT domain signature sequence. MutT is a protein involved in the GO system responsible for removing an oxidatively damaged form of guanine from DNA. A region of about 40 amino acid residues, found in the N-terminus of mutT, is also found in other proteins, including some phosphohydrolases (PROSITE PDOC00695).

Serine hydrolases are a large functional class of hydrolytic enzymes that contain a serine residue in their active site. This class of enzymes contains proteinases, esterases, and lipases which hydrolyze a variety of substrates and, therefore, have different biological roles. Proteins in this superfamily can be further grouped into subfamilies based on substrate specificity or amino acid similarities (Puente, X.S. and Lopez-Otin, C. (1995) J. Biol. Chem. 270: 12926-12932).

Neuropathy target esterase (NTE) is an integral membrane protein present in all neurons and in some non-neural-cell types of vertebrates. NTE is involved in a cell-signaling pathway controlling

interactions between neurons and accessory glial cells in the developing nervous system. NTE has serine esterase activity and efficiently catalyses the hydrolysis of phenyl valerate (PV) in vitro, but its physiological substrate is unknown. NTE is not related to either the major serine esterase family, which includes acetylcholinesterase, nor to any other known serine hydrolases. NTE contains at least two functional domains: an N-terminal putative regulatory domain and a C-terminal effector domain which contains the esterase activity and is, in part, conserved in proteins found in bacteria, yeast, nematodes and insects. NTE's effector domain contains three predicted transmembrane segments, and the active-site serine residue lies at the center of one of these segments. The isolated recombinant domain shows PV hydrolase activity only when incorporated into phospholipid liposomes. NTE's esterase activity is largely redundant in adult vertebrates, but organophosphates which react with NTE in vivo initiate unknown events which lead to a neuropathy with degeneration of long axons. These neuropathic organophosphates leave a negatively charged group covalently attached to the active-site serine residue, which causes a toxic gain of function in NTE (Glynn, P. (1999) Biochem. J. 344:625-631). Further, the Drosophila neurodegeneration gene swiss-cheese encodes a neuronal protein involved in glia-neuron interaction and is homologous to the above human NTE (Moser, M. et al. (2000) Mech. Dev. 90:279-282).

Chitinases are chitin-degrading enzymes present in a variety of organisms and participate in processes including cell wall remodeling, defense and catabolism. Chitinase activity has been found in human serum, leukocytes, granulocytes, and in association with fertilized oocytes in mammals (Escott, G.M. (1995) Infect. Immunol. 63:4770-4773; DeSouza, M.M. (1995) Endrocrinology 136:2485-2496). Glycolytic and proteolytic molecules in humans are associated with tissue damage in lung diseases and with increased tumorigenicity and metastatic potential of cancers (Mulligan, M.S. (1993) Proc. Natl. Acad. Sci. 90:11523-11527; Matrisian, L.M. (1991) Am. J. Med. Sci. 302:157-162; Witty, J.P. (1994) Cancer Res. 54:4805-4812). The discovery of a human enzyme with chitinolytic activity is noteworthy given the lack of endogenous chitin in the human body (Raghavan, N. (1994) Infect. Immun. 62:1901-1908). However, there is a group of mammalian proteins that share homology with chitinases from various non-mammalian organisms, such as bacteria, fungi, plants, and insects. The members of this family differ in their ability to hydrolyze chitin or chitin-like substrates. Some of the mammalian members of the family, such as a bovine whey chitotriosidase and human cartilage proteins which do not demonstrate specific chitinolytic activity, are expressed in association with tissue remodeling events (Rejman, J.J. (1988) Biochem. Biophys. Res. Commun. 150:329-334, Nyirkos, P. (1990) Biochem. J. 268:265-268). Elevated levels of human cartilage proteins have been reported in the synovial fluid and cartilage of patients with rheumatoid arthritis, a disease which produces a severe degradation of the cartilage and a proliferation of the synovial membrane in the affected joints (Hakala, B.E. (1993) J. Biol. Chem. 268:25803-25810).

A small subclass of hydrolases acting on ether bonds includes the thioether hydrolases. *S*-adenosyl-L-homocysteine hydrolase, also known as AdoHcyase or SAHH (PROSITE PDOC00603; EC 3.3.1.1), is a thioether hydrolase first described in rat liver extracts as the activity responsible for the reversible hydrolysis of *S*-adenosyl-L-homocysteine (AdoHcy) to adenosine and homocysteine (Sganga, M.W. et al. (1992) PNAS 89:6328-6332). SAHH is a cytosolic enzyme that has been found in all cells that have been tested, with the exception of Escherichia coli and certain related bacteria (Walker, R.D. et al. (1975) Can. J. Biochem. 53:312-319; Shimizu, S. et al. (1988) FEMS Microbiol. Lett. 51:177-180; Shimizu, S. et al. (1984) Eur. J. Biochem. 141:385-392). SAHH activity is dependent on $NAD^+$ as a cofactor. Deficiency of SAHH is associated with hypermethioninemia (Online Mendelian Inheritance in Man (OMIM) #180960 Hypermethioninemia), a pathologic condition characterized by neonatal cholestasis, failure to thrive, mental and motor retardation, facial dysmorphism with abnormal hair and teeth, and myocaridopathy (Labrune, P. et al. (1990) J. Pediat. 117:220-226).

Another subclass of hydrolases includes those enzymes which act on carbon-nitrogen (C-N) bonds other than peptide bonds. To this subclass belong those enzymes hydrolyzing amides, amidines, and other C-N bonds. This subclass is further subdivided on the basis of substrate specificity such as linear amides, cyclic amides, linear amidines, cyclic amidines, nitriles and other compounds. A hydrolase belonging to the sub-subclass of enzymes acting on the cyclic amidines is adenosine deaminase (ADA). ADA catalyzes the breakdown of adenosine to inosine. ADA is present in many mammalian tissues, including placenta, muscle, lung, stomach, digestive diverticulum, spleen, erythrocytes, thymus, seminal plasma, thyroid, T-cells, bone marrow stem cells, and liver. A subclass of ADAs, ADAR, act on RNA and are classified as RNA editases. An ADAR from Drosophila, dADAR, expressed in the developing nervous system, may act on para voltage-gated Na+ channel transcripts in the central nervous system (Palladino, M.J. et al. (2000) RNA 6:1004-1018). ADA deficiency causes profound lymphopenia with severe combined immunodeficiency (SCID). Cells from patients with ADA deficiency contain low, sometimes undetectable, amounts of ADA catalytic activity and ADA protein. ADA deficiency stems from genetic mutations in the ADA gene (Hershfield, M.S. (1998) Semin. Hematol. 4:291-298). Metabolic consequences of ADA deficiency are associated with defects in alveogenesis, pulmonary inflammation, and airway obstruction (Blackburn, M.R. et al. (2000) J. Exp. Med. 192:159-170).

Pancreatic ribonucleases (RNase) are pyrimidine-specific endonucleases found in high quantity in the pancreas of certain mammalian taxa and of some reptiles (Beintema, J.J. et al (1988) Prog. Biophys. Mol. Biol. 51:165-192). Proteins in the mammalian pancreatic RNase superfamily are noncytosolic endonucleases that degrade RNA through a two-step transphosphorolytic-hydrolytic reaction (Beintema, J.J. et al. (1986) Mol. Biol. Evol. 3:262-275). Specifically, the enzymes are

involved in endonucleolytic cleavage of 3'-phosphomononucleotides and 3'-phosphooligonucleotides ending in C-P or U-P with 2',3'-cyclic phosphate intermediates. Ribonucleases can unwind the DNA helix by complexing with single-stranded DNA; the complex arises by an extended multi-site cation-anion interaction between lysine and arginine residues of the enzyme and phosphate groups of the nucleotides. Some of the enzymes belonging to this family appear to play a purely digestive role, whereas others exhibit potent and unusual biological activities (D'Alessio, G. (1993) Trends Cell Biol. 3:106-109). Proteins belonging to the pancreatic RNase family include: bovine seminal vesicle and brain ribonucleases; kidney non-secretory ribonucleases (Beintema, J.J. et al (1986) FEBS Lett. 194:338-343); liver-type ribonucleases (Rosenberg, H.F. et al. (1989) PNAS U.S.A. 86:4460-4464); angiogenin, which induces vascularisation of normal and malignant tissues; eosinophil cationic protein (Hofsteenge, J. et al. (1989) Biochemistry 28:9806-9813), a cytotoxin and helminthotoxin with ribonuclease activity; and frog liver ribonuclease and frog sialic acid-binding lectin. The sequences of pancreatic RNases contain 4 conserved disulfide bonds and 3 amino acid residues involved in the catalytic activity.

ADP-ribosylation is a reversible post-translational protein modification in which an ADP-ribose moiety is transferred from β-NAD to a target amino acid such as arginine or cysteine. ADP-ribosylarginine hydrolases regenerate arginine by removing ADP-ribose from the protein, completing the ADP-ribosylation cycle (Moss, J. et al. (1997) Adv. Exp. Med. Biol. 419:25-33). ADP-ribosylation is a well-known reaction among bacterial toxins. Cholera toxin, for example, disrupts the adenylyl cyclase system by ADP-ribosylating the α-subunit of the stimulatory G-protein, causing an increase in intracellular cAMP (Moss, J. and Vaughan, M. (Eds) (1990) ADP-ribosylating Toxins and G-Proteins: Insights into Signal Transduction, American Society for Microbiology, Washington, D.C.). ADP-ribosylation may also have a regulatory function in eukaryotes, affecting such processes as cytoskeletal assembly (Zhou, H. et al. (1996) Arch. Biochem. Biophys. 334:214-222) and cell proliferation in cytotoxic T-cells (Wang, J. et al. (1996) J. Immunol. 156:2819-2827).

Nucleotidases catalyze the formation of free nucleosides from nucleotides. The cytosolic nucleotidase cN-I (5' nucleotidase-I) cloned from pigeon heart catalyzes the formation of adenosine from AMP generated during ATP hydrolysis (Sala-Newby, G.B. et al. (1999) J. Biol. Chem. 274:17789-17793). Increased adenosine concentration is thought to be a signal of metabolic stress, and adenosine receptors mediate effects including vasodilation, decreased stimulatory neuron firing and ischemic preconditioning in the heart (Schrader, J. (1990) Circulation 81:389-391; Rubino, A. et al. (1992) Eur. J. Pharmacol. 220:95-98; de Jong, J.W. et al. (2000) Pharmacol. Ther. 87:141-149). Deficiency of pyrimidine 5'-nucleotidase can result in hereditary hemolytic anemia (OMIM Entry 266120).

The lysozyme c superfamily consists of conventional lysozymes c, calcium-binding

lysozymes c, and α-lactalbumin (Prager, E.M. and P. Jolles (1996) EXS 75:9-31). The proteins in this superfamily have 35-40% sequence homology and share a common three-dimensional fold, but can have different functions. Lysozymes c are ubiquitous in a variety of tissues and secretions and can lyse the cell walls of certain bacteria (McKenzie, H.A. (1996) EXS 75:365-409). Alpha-lactalbumin is a metallo-protein that binds calcium and participates in the synthesis of lactose (Iyer, L.K. and P.K. Qasba (1999) Protein Eng. 12:129-139). Alpha-lactalbumin occurs in mammalian milk and colostrum (McKenzie, supra).

Lysozymes catalyze the hydrolysis of certain mucopolysaccharides of bacterial cell walls, specifically, the beta (1-4) glycosidic linkages between N-acetylmuramic acid and N-acetylglucosamine, and cause bacterial lysis. Lysozymes occur in diverse organisms including viruses, birds, and mammals. In humans, lysozymes are found in spleen, lung, kidney, white blood cells, plasma, saliva, milk, tears, and cartilage (Online Mendelian Inheritance in Man (OMIM) #153450 Lysozyme; Weaver, L.H. et al. (1985) J. Mol. Biol. 184:739-741). Lysozyme c functions in ruminants as a digestive enzyme, releasing proteins from ingested bacterial cells, and may perform the same function in human newborns (Braun, O.H. et al. (1995) Klin. Pediatr. 207:4-7).

The two known forms of lysozymes, chicken-type and goose-type, were originally isolated from chicken and goose egg white, respectively. Chicken-type and goose-type lysozymes have similar three-dimensional structures, but different amino acid sequences (Nakano, T. and T. Graf (1991) Biochim. Biophys. Acta 1090:273-276). In chickens, both forms of lysozyme are found in neutrophil granulocytes (heterophils), but only chicken-type lysozyme is found in egg white. Generally, chicken-type lysozyme mRNA is found in both adherent monocytes and macrophages and nonadherent promyelocytes and granulocytes as well as in cells of the bone marrow, spleen, bursa, and oviduct. Goose-type lysozyme mRNA is found in non-adherent cells of the bone marrow and lung. Several isozymes have been found in rabbits, including leukocytic, gastrointestinal, and possibly lymphoepithelial forms (OMIM #153450, supra; Nakano and Graf, supra; and GenBank GI 1310929). A human lysozyme gene encoding a protein similar to chicken-type lysozyme has been cloned (Yoshimura, K. et al. (1988) Biochem. Biophys. Res. Commun. 150:794-801). A consensus motif featuring regularly spaced cysteine residues has been derived from the lysozyme C enzymes of various species (PROSITE PS00128). Lysozyme C shares about 40% amino acid sequence identity with α-lactalbumin.

Lysozymes have several disease associations. Lysozymuria is observed in diabetic nephropathy (Shima, M. et al.(1986) Clin. Chem. 32:1818-1822), endemic nephropathy (Bruckner, I. et al. (1978) Med. Interne. 16:117-125), urinary tract infections (Heidegger, H. (1990) Minerva Ginecol. 42:243-250), and acute monocytic leukemia (Shaw, M.T. (1978) Am. J. Hematol. 4:97-103). Nakano (supra) suggested a role for lysozyme in host defense systems. Older rabbits with an

inherited lysozyme deficiency show increased susceptibility to infections, such as subcutaneous abscesses (OMIM #153450, supra). Human lysozyme gene mutations cause hereditary systemic amyloidosis, a rare autosomal dominant disease in which amyloid deposits form in the viscera, including the kidney, adrenal glands, spleen, and liver. This disease is usually fatal by the fifth decade. The amyloid deposits contain variant forms of lysozyme. Renal amyloidosis is the most common and potentially the most serious form of organ involvement (Pepys, M.B. et al. (1993) Nature 362:553-557; OMIM #105200 Familial Visceral Amyloidosis; Cotran, R.S. et al. (1994) Robbins Pathologic Basis of Disease, W.B. Saunders Company, Philadelphia PA, pp. 231-238). Increased levels of lysozyme and lactate have been observed in the cerebrospinal fluid of patients with bacterial meningitis (Ponka, A. et al. (1983) Infection 11:129-131). Acute monocytic leukemia is characterized by massive lysozymuria (Den Tandt, W.R. (1988) Int. J. Biochem. 20:713-719).

Lyases

Lyases are a class of enzymes that catalyze the cleavage of C-C, C-O, C-N, C-S, C-(halide), P-O, or other bonds without hydrolysis or oxidation to form two molecules, at least one of which contains a double bond (Stryer, L. (1995) Biochemistry, W.H. Freeman and Co., New York NY, p.620). Under the International Classification of Enzymes (Webb, E. C. (1992) Enzyme Nomenclature 1992: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes, Academic Press, San Diego CA), lyases form a distinct class designated by the numeral 4 in the first digit of the enzyme number (i.e., EC 4.x.x.x).

Further classification of lyases reflects the type of bond cleaved as well as the nature of the cleaved group. The group of C-C lyases includes carboxyl-lyases (decarboxylases), aldehyde-lyases (aldolases), oxo-acid-lyases, and other lyases. The C-O lyase group includes hydro-lyases, lyases acting on polysaccharides, and other lyases. The C-N lyase group includes ammonia-lyases, amidine-lyases, amine-lyases (deaminases), and other lyases. Lyases are critical components of cellular biochemistry, with roles in metabolic energy production, including fatty acid metabolism and the tricarboxylic acid cycle, as well as other diverse enzymatic processes.

One important family of lyases are the carbonic anhydrases (CA), also called carbonate dehydratases, which catalyze the hydration of carbon dioxide in the reaction $H_2O + CO_2 \rightleftharpoons HCO_3^- + H^+$. CA accelerates this reaction by a factor of over $10^6$ by virtue of a zinc ion located in a deep cleft about 15Å below the protein's surface and co-ordinated to the imidazole groups of three His residues. Water bound to the zinc ion is rapidly converted to $HCO_3^-$.

Eight enzymatic and evolutionarily related forms of carbonic anhydrase are currently known to exist in humans: three cytosolic isozymes (CAI, CAII, and CAIII), two membrane-bound forms

(CAIV and CAVII), a mitochondrial form (CAV), a secreted salivary form (CAVI) and a yet uncharacterized isozyme (PROSITE PDOC00146 Eukaryotic-type carbonic anhydrases signature). Though the isoenzymes CAI, CAII, and bovine CAIII have similar secondary structures and polypeptide-chain folds, CAI has 6 tryptophans, CAII has 7 and CAIII has 8 (Boren, K. et al. (1996) Protein Sci. 5:2479-2484). CAII is the predominant CA isoenzyme in the brain of mammals.

CAs participate in a variety of physiological processes that involve pH regulation, $CO_2$ and $HCO_3^-$ transport, ion transport, and water and electrolyte balance. For example, CAII contributes to $H^+$ secretion by gastric parietal cells, by renal tubular cells, and by osteoclasts that secrete $H^+$ to acidify the bone-resorbing compartment. In addition, CAII promotes $HCO_3^-$ secretion by pancreatic duct cells, cilary body epithelium, choroid plexus, salivary gland acinar cells, and distal colonal epithelium, thus playing a role in the production of pancreatic juice, aqueous humor, cerebrospinal fluid, and saliva, and contributing to electrolyte and water balance. CAII also promotes $CO_2$ exchange in proximal tubules in the kidney, in erythrocytes, and in lung. CAIV has roles in several tissues: it facilitates $HCO_3^-$ reabsorption in the kidney; promotes $CO_2$ flux in tissues including brain, skeletal muscle, and heart muscle; and promotes $CO_2$ exchange from the blood to the alveoli in the lung. CAVI probably plays a role in pH regulation in saliva, along with CAII, and may have a protective effect in the esophagus and stomach. Mitochondrial CAV appears to play important roles in gluconeogenesis and ureagenesis, based on the effects of CA inhibitors on these pathways. (Sly, W.S. and Hu, P.Y. (1995) Ann. Rev. Biochem. 64:375-401.)

A number of disease states are marked by variations in CA activity. Mutations in CAII which lead to CAII deficiency are the cause of osteopetrosis with renal tubular acidosis (Online Medelian Inheritance in Man 259730 Osteopetrosis with Renal Tubular Acidosis). The concentration of CAII in the cerebrospinal fluid (CSF) appears to mark disease activity in patients with brain damage. High CA concentrations have been observed in patients with brain infarction. Patients with transient ischemic attack, multiple sclerosis, or epilepsy usually have CAII concentrations in the normal range, but higher CAII levels have been observed in the CSF of those with central nervous system infection, dementia, or trigeminal neuralgia (Parkkila, A.K. et al. (1997) Eur. J. Clin. Invest. 27:392-397). Colonic adenomas and adenocarcinomas have been observed to fail to stain for CA, whereas non-neoplastic controls showed CAI and CAII in the cytoplasm of the columnar cells lining the upper half of colonic crypts. The neoplasms show staining patterns similar to less mature cells lining the base of normal crypts (Gramlich T.L. et al. (1990) Arch. Pathol. Lab. Med. 114:415-419).

Therapeutic interventions in a number of diseases involve altering CA activity. CA inhibitors such as acetazolamide are used in the treatment of glaucoma (Stewart, W.C. (1999) Curr. Opin. Opthamol. 10:99-108), essential tremor and Parkinson's disease (Uitti, R.J. (1998) Geriatrics 53:46-48, 53-57), intermittent ataxia (Singhvi, J.P. et al. (2000) Neurology India 48:78-80), and altitude

related illnesses (Klocke, D.L. et al. (1998) Mayo Clin. Proc. 73:988-992).

CA activity can be particularly useful as an indicator of long-term disease conditions, since the enzyme reacts relatively slowly to physiological changes. CAI and zinc concentrations have been observed to decrease in hyperthyroid Graves' disease (Yoshida, K. (1996) Tohoku J. Exp. Med. 178:345-356) and glycosylated CAI is observed in diabetes mellitus (Kondo, T. et al. (1987) Clin. Chim. Acta 166:227-236). A positive correlation has been observed between CAI and CAII reactivity and endometriosis (Brinton, D.A. et al. (1996) Ann. Clin. Lab. Sci. 26:409-420; D'Cruz , O.J. et al. (1996) Fertil. Steril. 66:547-556).

Another important member of the lyase family is ornithine decarboxylase (ODC), the initial rate-limiting enzyme in polyamine biosynthesis. ODC catalyses the transformation of ornithine into putrescine in the reaction L-ornithine $\rightleftharpoons$ putrescine + $CO_2$. Polyamines, which include putrescine and the subsequent metabolic pathway products spermidine and spermine, are ubiquitous cell components essential for DNA synthesis, cell differentiation, and proliferation. Thus the polyamines play a key role in tumor proliferation (Medina, M.A. et al. (1999) Biochem. Pharmacol. 57:1341-1344).

ODC is a pyridoxal-5'-phosphate (PLP)-dependent enzyme which is active as a homodimer. Conserved residues include those at the PLP binding site and a stretch of glycine residues thought to be part of a substrate binding region (PROSITE PDOC00685 Orn/DAP/Arg decarboxylase family 2 signatures). Mammalian ODCs also contain PEST regions, sequence fragments enriched in proline, glutamic acid, serine, and threonine residues that act as signals for intracellular degradation (Medina, supra).

Many chemical carcinogens and tumor promoters increase ODC levels and activity. Several known oncogenes may increase ODC levels by enhancing transcription of the ODC gene, and ODC itself may act as an oncogene when expressed at very high levels. A high level of ODC is found in a number of precancerous conditions, and elevation of ODC levels has been used as part of a screen for tumor-promoting compounds (Pegg, A.E. et al. (1995) J. Cell. Biochem. Suppl. 22:132-138).

Inhibitors of ODC have been used to treat tumors in animal models and human clinical trials, and have been shown to reduce development of tumors of the bladder, brain, esophagus, gastrointestinal tract, lung, oral cavity, mammary gland, stomach, skin and trachea (Pegg, supra; McCann, P.P. and A.E. Pegg (1992) Pharmac. Ther. 54:195-215). ODC also shows promise as a target for chemoprevention (Pegg, supra). ODC inhibitors have also been used to treat infections by African trypanosomes, malaria, and Pneumocystis carinii, and are potentially useful for treatment of autoimmune diseases such as lupus and rheumatoid arthritis (McCann, supra).

Another family of pyridoxal-dependent decarboxylases are the group II decarboxylases. This family includes glutamate decarboxylase (GAD) which catalyzes the decarboxylation of glutamate into the neurotransmitter GABA; histidine decarboxylase (HDC), which catalyzes the

decarboxylation of histidine to histamine; aromatic-L-amino-acid decarboxylase (DDC), also known as L-dopa decarboxylase or tryptophan decarboxylase, which catalyzes the decarboxylation of tryptophan to tryptamine and also acts on 5-hydroxy-tryptophan and dihydroxyphenylalanine (L-dopa); and cysteine sulfinic acid decarboxylase (CSD), the rate-limiting enzyme in the synthesis of taurine from cysteine (PROSITE PDOC00329 DDC/GAD/HDC/TyrDC pyridoxal-phosphate attachment site). Taurine is an abundant sulfonic amino acid in brain and is thought to act as an osmoregulator in brain cells (Bitoun, M. and Tappaz, M. (2000) J. Neurochem. 75:919-924).

Isomerases

Isomerases are a class of enzymes that catalyze geometric or structural changes within a molecule to form a single product. This class includes racemases and epimerases, cis-trans-isomerases, intramolecular oxidoreductases, intramolecular transferases (mutases) and intramolecular lyases. Isomerases are critical components of cellular biochemistry with roles in metabolic energy production including glycolysis, as well as other diverse enzymatic processes (Stryer, L. (1995) Biochemistry W.H. Freeman and Co. New York, NY pp.483-507).

Racemases are a subset of isomerases that catalyze inversion of a molecule's configuration around the asymmetric carbon atom in a substrate having a single center of asymmetry, thereby interconverting two racemers. Epimerases are another subset of isomerases that catalyze inversion of configuration around an asymmetric carbon atom in a substrate with more than one center of symmetry, thereby interconverting two epimers. Racemases and epimerases can act on amino acids and derivatives, hydroxy acids and derivatives, and carbohydrates and derivatives. The interconversion of UDP-galactose and UDP-glucose is catalyzed by UDP-galactose-4'-epimerase. Proper regulation and function of this epimerase is essential to the synthesis of glycoproteins and glycolipids. Elevated blood galactose levels have been correlated with UDP-galactose-4'-epimerase deficiency in screening programs of infants (Gitzelmann, R. (1972) Helv. Paediat. Acta 27:125-130).

Correct folding of newly synthesized proteins is assisted by molecular chaperones and folding catalysts, two unrelated groups of helper molecules. Chaperones suppress non-productive side reactions by stoichiometric binding to folding intermediates, whereas folding enzymes catalyze some of the multiple folding steps that enable proteins to attain their final functional configurations (Kern, G. et al. (1994) FEBS Lett. 348: 145-148). One class of folding enzymes, the peptidyl prolyl *cis-trans* isomerases (PPIases), isomerizes certain proline imidic bonds in what is considered to be a rate limiting step in protein maturation and export. PPIases catalyze the *cis* to *trans* isomerization of certain proline imidic bonds in proteins. There are three evolutionarily unrelated families of PPIases: the cyclophilins, the FK506 binding proteins, and the newly characterized parvulin family (Rahfeld, J.U. et al. (1994) FEBS Lett. 352: 180-184).

The cyclophilins (CyP) were originally identified as major receptors for the immunosuppressive drug cyclosporin A (CsA), an inhibitor of T-cell activation (Handschumacher, R.E. et al. (1984) Science 226: 544-547; Harding, M.W. et al. (1986) J. Biol. Chem. 261: 8547-8555). Thus, the peptidyl-prolyl isomerase activity of CyP may be part of the signaling pathway that leads to T-cell activation. Subsequent work demonstrated that CyP's isomerase activity is essential for correct protein folding and/or protein trafficking, and may also be involved in assembly/disassembly of protein complexes and regulation of protein activity. For example, in *Drosophila*, the CyP NinaA is required for correct localization of rhodopsins, while a mammalian CyP (Cyp40) is part of the Hsp90/Hsp70 complex that binds steroid receptors. The mammalian CyP (CypA) has been shown to bind the *gag* protein from human immunodeficiency virus 1 (HIV-1), an interaction that can be inhibited by cyclosporin. Since cyclosporin has potent anti-HIV-1 activity, CypA may play an essential function in HIV-1 replication. Finally, Cyp40 has been shown to bind and inactivate the transcription factor c-Myb, an effect that is reversed by cyclosporin. This effect implicates CyP in the regulation of transcription, transformation, and differentiation (Bergsma, D.J. et al (1991) J. Biol. Chem. 266:23204-23214; Hunter, T. (1998) Cell 92: 141-143; and Leverson, J.D. and Ness, S.A. (1998) Mol. Cell. 1:203-211).

One of the major rate limiting steps in protein folding is the thiol:disulfide exchange that is necessary for correct protein assembly. Although incubation of reduced, unfolded proteins in buffers with defined ratios of oxidized and reduced thiols can lead to native conformation, the rate of folding is slow and the attainment of native conformation decreases proportionately with the size and number of cysteines in the protein. Certain cellular compartments such as the endoplasmic reticulum of eukaryotes and the periplasmic space of prokaryotes are maintained in a more oxidized state than the surrounding cytosol. Correct disulfide formation can occur in these compartments, but at a rate that is insufficient for normal cell processes and inadequate for synthesizing secreted proteins. The protein disulfide isomerases, thioredoxins and glutaredoxins are able to catalyze the formation of disulfide bonds and regulate the redox environment in cells to enable the necessary thiol:disulfide exchanges (Loferer, H. (1995) J. Biol. Chem. 270:26178-26183).

Each of these proteins has somewhat different functions, but all belong to a group of disulfide-containing redox proteins that contain a conserved active-site sequence and are ubiquitously distributed in eukaryotes and prokaryotes. Protein disulfide isomerases are found in the endoplasmic reticulum of eukaryotes and in the periplasmic space of prokaryotes. They function by exchanging their own disulfide for a thiol in a folding peptide chain. In contrast, the reduced thioredoxins and glutaredoxins are generally found in the cytoplasm and function by directly reducing disulfides in the substrate proteins.

Oxidoreductases can be isomerases as well. Oxidoreductases catalyze the reversible transfer

of electrons from a substrate that becomes oxidized to a substrate that becomes reduced. This class of enzymes includes dehydrogenases, hydroxylases, oxidases, oxygenases, peroxidases, and reductases. Proper maintenance of oxidoreductase levels is physiologically important. For example, genetically-linked deficiencies in lipoamide dehydrogenase can result in lactic acidosis (Robinson, B. H. et. al. (1977) Pediat. Res. 11:1198-1202).

Another subgroup of isomerases are the transferases (or mutases). Transferases transfer a chemical group from one compound (the donor) to another compound (the acceptor). The types of groups transferred by these enzymes include acyl groups, amino groups, phosphate groups (phosphotransferases or phosphomutases), and others. The transferase carnitine palmitoyltransferase is an important component of fatty acid metabolism. Genetically-linked deficiencies in this transferase can lead to myopathy (Scriver C. . et. al. (1995) The Metabolic and Molecular Basis of Inherited Disease, McGraw-Hill New York NY pp.1501-1533).

Yet another subgroup of isomerases are the topoisomersases. Topoisomerases are enzymes that affect the topological state of DNA. For example, defects in topoisomerases or their regulation can affect normal physiology. Reduced levels of topoisomerase II have been correlated with some of the DNA processing defects associated with the disorder ataxia-telangiectasia (Singh, S.P. et. al. (1988) Nucleic Acids Res. 16:3919-3929).

Ligases

Ligases catalyze the formation of a bond between two substrate molecules. The process involves the hydrolysis of a pyrophosphate bond in ATP or a similar energy donor. Ligases are classified based on the nature of the type of bond they form, which can include carbon-oxygen, carbon-sulfur, carbon-nitrogen, carbon-carbon and phosphoric ester bonds.

Ligases forming carbon-oxygen bonds include the aminoacyl-transfer RNA (tRNA) synthetases which are important RNA-associated enzymes with roles in translation. Protein biosynthesis depends on each amino acid forming a linkage with the appropriate tRNA. The aminoacyl-tRNA synthetases are responsible for the activation and correct attachment of an amino acid with its cognate tRNA. The 20 aminoacyl-tRNA synthetase enzymes can be divided into two structural classes, and each class is characterized by a distinctive topology of the catalytic domain. Class I enzymes contain a catalytic domain based on the nucleotide-binding "Rossman fold". Class II enzymes contain a central catalytic domain, which consists of a seven-stranded antiparallel ß-sheet motif, as well as N- and C- terminal regulatory domains. Class II enzymes are separated into two groups based on the heterodimeric or homodimeric structure of the enzyme; the latter group is further subdivided by the structure of the N- and C-terminal regulatory domains (Hartlein, M. and Cusack, S. (1995) J. Mol. Evol. 40:519-530). Autoantibodies against aminoacyl-tRNAs are generated by

patients with dermatomyositis and polymyositis, and correlate strongly with complicating interstitial lung disease (ILD). These antibodies appear to be generated in response to viral infection, and coxsackie virus has been used to induce experimental viral myositis in animals.

Ligases forming carbon-sulfur bonds (acid-thiol ligases) mediate a large number of cellular biosynthetic intermediary metabolism processes involving intermolecular transfer of carbon atom-containing substrates (carbon substrates). Examples of such reactions include the tricarboxylic acid cycle, synthesis of fatty acids and long-chain phospholipids, synthesis of alcohols and aldehydes, synthesis of intermediary metabolites, and reactions involved in the amino acid degradation pathways. Some of these reactions require input of energy, usually in the form of conversion of ATP to either ADP or AMP and pyrophosphate.

In many cases, a carbon substrate is derived from a small molecule containing at least two carbon atoms. The carbon substrate is often covalently bound to a larger molecule which acts as a carbon substrate carrier molecule within the cell. In the biosynthetic mechanisms described above, the carrier molecule is coenzyme A. Coenzyme A (CoA) is structurally related to derivatives of the nucleotide ADP and consists of 4'-phosphopantetheine linked via a phosphodiester bond to the alpha phosphate group of adenosine 3',5'-bisphosphate. The terminal thiol group of 4'-phosphopantetheine acts as the site for carbon substrate bond formation. The predominant carbon substrates which utilize CoA as a carrier molecule during biosynthesis and intermediary metabolism in the cell are acetyl, succinyl, and propionyl moieties, collectively referred to as acyl groups. Other carbon substrates include enoyl lipid, which acts as a fatty acid oxidation intermediate, and carnitine, which acts as an acetyl-CoA flux regulator/mitochondrial acyl group transfer protein. Acyl-CoA and acetyl-CoA are synthesized in the cell by acyl-CoA synthetase and acetyl-CoA synthetase, respectively.

Activation of fatty acids is mediated by at least three forms of acyl-CoA synthetase activity: i) acetyl-CoA synthetase, which activates acetate and several other low molecular weight carboxylic acids and is found in muscle mitochondria and the cytosol of other tissues; ii) medium-chain acyl-CoA synthetase, which activates fatty acids containing between four and eleven carbon atoms (predominantly from dietary sources), and is present only in liver mitochondria; and iii) acyl CoA synthetase, which is specific for long chain fatty acids with between six and twenty carbon atoms, and is found in microsomes and the mitochondria. Proteins associated with acyl-CoA synthetase activity have been identified from many sources including bacteria, yeast, plants, mouse, and man. The activity of acyl-CoA synthetase may be modulated by phosphorylation of the enzyme by cAMP-dependent protein kinase.

Ligases forming carbon-nitrogen bonds include amide synthases such as glutamine synthetase (glutamate-ammonia ligase) that catalyzes the amination of glutamic acid to glutamine by ammonia using the energy of ATP hydrolysis. Glutamine is the primary source for the amino group

in various amide transfer reactions involved in de novo pyrimidine nucleotide synthesis and in purine and pyrimidine ribonucleotide interconversions. Overexpression of glutamine synthetase has been observed in primary liver cancer (Christa, L. et al. (1994) Gastroent. 106:1312-1320).

Acid-amino-acid ligases (peptide synthases) are represented by the ubiquitin conjugating enzymes which are associated with the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins in eukaryotic cells and some bacteria. The UCS mediates the elimination of abnormal proteins and regulates the half-lives of important regulatory proteins that control cellular processes such as gene transcription and cell cycle progression. In the UCS pathway, proteins targeted for degradation are conjugated to ubiquitin (Ub), a small heat stable protein. Ub is first activated by a ubiquitin-activating enzyme (E1), and then transferred to one of several Ub-conjugating enzymes (E2). E2 then links the Ub molecule through its C-terminal glycine to an internal lysine (acceptor lysine) of a target protein. The ubiquitinated protein is then recognized and degraded by proteasome, a large, multisubunit proteolytic enzyme complex, and ubiquitin is released for reutilization by ubiquitin protease. The UCS is implicated in the degradation of mitotic cyclic kinases, oncoproteins, tumor suppressor genes such as p53, viral proteins, cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, A. (1994) Cell 79:13-21).

Cyclo-ligases and other carbon-nitrogen ligases comprise various enzymes and enzyme complexes that participate in the de novo pathways of purine and pyrimidine biosynthesis. Because these pathways are critical to the synthesis of nucleotides for replication of both RNA and DNA, many of these enzymes have been the targets of clinical agents for the treatment of cell proliferative disorders such as cancer and infectious diseases.

Purine biosynthesis occurs de novo from the amino acids glycine and glutamine, and other small molecules. Three of the key reactions in this process are catalyzed by a trifunctional enzyme composed of glycinamide-ribonucleotide synthetase (GARS), aminoimidazole ribonucleotide synthetase (AIRS), and glycinamide ribonucleotide transformylase (GART). Together these three enzymes combine ribosylamine phosphate with glycine to yield phosphoribosyl aminoimidazole, a precursor to both adenylate and guanylate nucleotides. This trifunctional protein has been implicated in the pathology of Downs syndrome (Aimi, J. et al. (1990) Nucleic Acid Res. 18:6665-6672). Adenylosuccinate synthetase catalyzes a later step in purine biosynthesis that converts inosinic acid to adenylosuccinate, a key step on the path to ATP synthesis. This enzyme is also similar to another carbon-nitrogen ligase, argininosuccinate synthetase, that catalyzes a similar reaction in the urea cycle (Powell, S.M. et al. (1992) FEBS Lett. 303:4-10).

Adenylosuccinate synthetase, adenylosuccinate lyase, and AMP deaminase may be considered as a functional unit, the purine nucleotide cycle. This cycle converts AMP to inosine

monophosphate (IMP) and reconverts IMP to AMP via adenylosuccinate, thereby producing $NH_3$ and forming fumarate from aspartate. In muscle, the purine nucleotide cycle functions, during intense exercise, in the regeneration of ATP by pulling the adenylate kinase reaction in the direction of ATP formation and by providing Krebs cycle intermediates. In kidney, the purine nucleotide cycle accounts for the release of $NH_3$ under normal acid-base conditions. In brain, the purine nucleotide cycle may contribute to ATP recovery. Adenylosuccinate lyase deficiency provokes psychomotor retardation, often accompanied by autistic features (Van den Berghe, G. et al. (1992) Prog Neurobiol.: 39:547-561). A marked imbalance in the enzymic pattern of purine metabolism is linked with transformation and/or progression in cancer cells. In rat hepatomas the specific activities of the anabolic enzymes, IMP dehydrogenase, GMP synthetase, adenylosuccinate synthetase, adenylosuccinase, AMP deaminase and amidophosphoribosyltransferase, increased to 13.5-, 3.7-, 3.1-, 1.8-, 5.5- and 2.8-fold, respectively, of those in normal liver (Weber, G. (1983) Clin Biochem 1983 Feb;16(1):57-63).

Like the de novo biosynthesis of purines, de novo synthesis of the pyrimidine nucleotides uridylate and cytidylate also arises from a common precursor, in this instance the nucleotide orotidylate derived from orotate and phosphoribosyl pyrophosphate (PPRP). Again a trifunctional enzyme comprising three carbon-nitrogen ligases plays a key role in the process. In this case the enzymes aspartate transcarbamylase (ATCase), carbamyl phosphate synthetase II, and dihydroorotase (DHOase) are encoded by a single gene called CAD. Together these three enzymes combine the initial reactants in pyrimidine biosynthesis, glutamine, $CO_2$, and ATP to form dihydroorotate, the precursor to orotate and orotidylate (Iwahana, H. et al. (1996) Biochem. Biophys. Res. Commun. 219:249-255). Further steps then lead to the synthesis of uridine nucleotides from orotidylate. Cytidine nucleotides are derived from uridine-5'-triphosphate (UTP) by the amidation of UTP using glutamine as the amino donor and the enzyme CTP synthetase. Regulatory mutations in the human CTP synthetase are believed to confer multi-drug resistance to agents widely used in cancer therapy (Yamauchi, M. et al. (1990) EMBO J. 9:2095-2099).

Ligases forming carbon-carbon bonds include the carboxylases acetyl-CoA carboxylase and pyruvate carboxylase. Acetyl-CoA carboxylase catalyzes the carboxylation of acetyl-CoA from $CO_2$ and $H_2O$ using the energy of ATP hydrolysis. Acetyl-CoA carboxylase is the rate-limiting enzyme in the biogenesis of long-chain fatty acids. Two isoforms of acetyl-CoA carboxylase, types I and types II, are expressed in human in a tissue-specific manner (Ha, J. et al. (1994) Eur. J. Biochem. 219:297-306). Pyruvate carboxylase is a nuclear-encoded mitochondrial enzyme that catalyzes the conversion of pyruvate to oxaloacetate, a key intermediate in the citric acid cycle.

Ligases forming phosphoric ester bonds include the DNA ligases involved in both DNA replication and repair. DNA ligases seal phosphodiester bonds between two adjacent nucleotides in a

DNA chain using the energy from ATP hydrolysis to first activate the free 5'-phosphate of one nucleotide and then react it with the 3'-OH group of the adjacent nucleotide. This resealing reaction is used in DNA replication to join small DNA fragments called "Okazaki" fragments that are transiently formed in the process of replicating new DNA, and in DNA repair. DNA repair is the process by which accidental base changes, such as those produced by oxidative damage, hydrolytic attack, or uncontrolled methylation of DNA, are corrected before replication or transcription of the DNA can occur. Bloom's syndrome is an inherited human disease in which individuals are partially deficient in DNA ligation and consequently have an increased incidence of cancer (Alberts, B. et al. (1994) The Molecular Biology of the Cell, Garland Publishing Inc., New York, NY, p. 247).

Pantothenate synthetase (D-pantoate; beta-alanine ligase (AMP-forming); EC 6.3.2.1) is the last enzyme of the pathway of pantothenate (vitamin B(5)) synthesis. It catalyzes the condensation of pantoate with beta-alanine in an ATP-dependent reaction. The enzyme is dimeric, with two well-defined domains per protomer: the N-terminal domain, a Rossmann fold, contains the active site cavity, with the C-terminal domain forming a hinged lid. The N-terminal domain is structurally very similar to class I aminoacyl-tRNA synthetases and is thus a member of the cytidylyltransferase superfamily (von Delft, F. et al. (2000) Structure (Camb) 9:439-450).

Farnesyl diphosphate synthase (FPPS) is an essential enzyme that is required both for cholesterol synthesis and protein prenylation. The enzyme catalyzes the formation of farnesyl diphosphate from dimethylallyl diphosphate and isopentyl diphosphate. FPPS is inhibited by nitrogen-containing biphosphonates, which can lead to the inhibition of osteoclast-mediated bone resorption by preventing protein prenylation (Dunford, J.E. et al. (2001) J. Pharmacol. Exp. Ther. 296:235-242).

5-aminolevulinate synthase (ALAS; delta-aminolevulinate synthase; EC 2.3.1.37) catalyzes the rate-limiting step in heme biosynthesis in both erythroid and non-erythroid tissues. This enzyme is unique in the heme biosynthetic pathway in being encoded by two genes, the first encoding ALAS1, the non-erythroid specific enzyme which is ubiquitously expressed, and the second encoding ALAS2, which is expressed exclusively in erythroid cells. The genes for ALAS1 and ALAS2 are located, respectively, on chromosome 3 and on the X chromosome. Defects in the gene encoding ALAS2 result in X-linked sideroblastic anemia. Elevated levels of ALAS are seen in acute hepatic porphyrias and can be lowered by zinc mesoporphyrin.

Drug Metabolizing Enzymes (DMEs)

The metabolism of a drug and its movement through the body (pharmacokinetics) are important in determining its effects, toxicity, and interactions with other drugs. The three processes governing pharmacokinetics are the absorption of the drug, distribution to various tissues, and

elimination of drug metabolites. These processes are intimately coupled to drug metabolism, since a variety of metabolic modifications alter most of the physicochemical and pharmacological properties of drugs, including solubility, binding to receptors, and excretion rates. The metabolic pathways which modify drugs also accept a variety of naturally occurring substrates such as steroids, fatty acids, prostaglandins, leukotrienes, and vitamins. The enzymes in these pathways are therefore important sites of biochemical and pharmacological interaction between natural compounds, drugs, carcinogens, mutagens, and xenobiotics. It has long been appreciated that inherited differences in drug metabolism lead to drastically different levels of drug efficacy and toxicity among individuals. Advances in pharmacogenomics research, of which DMEs constitute an important part, are promising to expand the tools and information that can be brought to bear on questions of drug efficacy and toxicity (See Evans, W.E. and R.V. Relling (1999) Science 286:487-491). DMEs have broad substrate specificities, unlike antibodies, for example, which are diverse and highly specific. Since DMEs metabolize a wide variety of molecules, drug interactions may occur at the level of metabolism so that, for example, one compound may induce a DME that affects the metabolism of another compound.

Drug metabolic reactions are categorized as Phase I, which prepare the drug molecule for functioning and further metabolism, and Phase II, which are conjugative. In general, Phase I reaction products are partially or fully inactive, and Phase II reaction products are the chief excreted species. However, Phase I reaction products are sometimes more active than the original administered drugs; this metabolic activation principle is exploited by pro-drugs (e.g. L-dopa). Additionally, some nontoxic compounds (e.g. aflatoxin, benzo[*a*]pyrene) are metabolized to toxic intermediates through these pathways. Phase I reactions are usually rate-limiting in drug metabolism. Prior exposure to the compound, or other compounds, can induce the expression of Phase I enzymes however, and thereby increase substrate flux through the metabolic pathways. (See Klaassen, C.D. et al. (1996) Casarett and Doull's Toxicology: The Basic Science of Poisons, McGraw-Hill, New York, NY, pp. 113-186; Katzung, B.G. (1995) Basic and Clinical Pharmacology, Appleton and Lange, Norwalk, CT, pp. 48-59; Gibson, G.G. and P. Skett (1994) Introduction to Drug Metabolism, Blackie Academic and Professional, London.).

The major classes of Phase I enzymes include, but are not limited to, cytochrome P450 and flavin-containing monooxygenase. Other enzyme classes involved in Phase I-type catalytic cycles and reactions include, but are not limited to, NADPH cytochrome P450 reductase (CPR), the microsomal cytochrome b5/NADH cytochrome b5 reductase system, the ferredoxin/ferredoxin reductase redox pair, aldo/keto reductases, and alcohol dehydrogenases. The major classes of Phase II enzymes include, but are not limited to, UDP glucuronyltransferase, sulfotransferase, glutathione S-transferase, N-acyltransferase, and N-acetyl transferase.

Cytochrome P450 and P450 catalytic cycle-associated enzymes

Members of the cytochrome P450 superfamily of enzymes catalyze the oxidative metabolism of a variety of substrates, including natural compounds such as steroids, fatty acids, prostaglandins, leukotrienes, and vitamins, as well as drugs, carcinogens, mutagens, and xenobiotics. Cytochromes P450, also known as P450 heme-thiolate proteins, usually act as terminal oxidases in multi-component electron transfer chains, called P450-containing monooxygenase systems. Specific reactions catalyzed include hydroxylation, epoxidation, N-oxidation, sulfooxidation, N-, S-, and O-dealkylations, desulfation, deamination, and reduction of azo, nitro, and N-oxide groups. These reactions are involved in steroidogenesis of glucocorticoids, cortisols, estrogens, and androgens in animals; insecticide resistance in insects; herbicide resistance and flower coloring in plants; and environmental bioremediation by microorganisms. Cytochrome P450 actions on drugs, carcinogens, mutagens, and xenobiotics can result in detoxification or in conversion of the substance to a more toxic product. Cytochromes P450 are abundant in the liver, but also occur in other tissues; the enzymes are located in microsomes. (See ExPASY ENZYME EC 1.14.14.1; Prosite PDOC00081 Cytochrome P450 cysteine heme-iron ligand signature; PRINTS EP450I E-Class P450 Group I signature; Graham-Lorence, S. and Peterson, J.A. (1996) FASEB J. 10:206-214.)

Four hundred cytochromes P450 have been identified in diverse organisms including bacteria, fungi, plants, and animals (Graham-Lorence, supra). The B-class is found in prokaryotes and fungi, while the E-class is found in bacteria, plants, insects, vertebrates, and mammals. Five subclasses or groups are found within the larger family of E-class cytochromes P450 (PRINTS EP450I E-Class P450 Group I signature).

All cytochromes P450 use a heme cofactor and share structural attributes. Most cytochromes P450 are 400 to 530 amino acids in length. The secondary structure of the enzyme is about 70% alpha-helical and about 22% beta-sheet. The region around the heme-binding site in the C-terminal part of the protein is conserved among cytochromes P450. A ten amino acid signature sequence in this heme-iron ligand region has been identified which includes a conserved cysteine involved in binding the heme iron in the fifth coordination site. In eukaryotic cytochromes P450, a membrane-spanning region is usually found in the first 15-20 amino acids of the protein, generally consisting of approximately 15 hydrophobic residues followed by a positively charged residue. (See Prosite PDOC00081, supra; Graham-Lorence, supra.)

Cytochrome P450 enzymes are involved in cell proliferation and development. The enzymes have roles in chemical mutagenesis and carcinogenesis by metabolizing chemicals to reactive intermediates that form adducts with DNA (Nebert, D.W. and Gonzalez, F.J. (1987) Ann. Rev. Biochem. 56:945-993). These adducts can cause nucleotide changes and DNA rearrangements that lead to oncogenesis. Cytochrome P450 expression in liver and other tissues is induced by xenobiotics

such as polycyclic aromatic hydrocarbons, peroxisomal proliferators, phenobarbital, and the glucocorticoid dexamethasone (Dogra, S.C. et al. (1998) Clin. Exp. Pharmacol. Physiol. 25:1-9). A cytochrome P450 protein may participate in eye development as mutations in the P450 gene CYP1B1 cause primary congenital glaucoma (Online Mendelian Inheritance in Man (OMIM) *601771 Cytochrome P450, subfamily I (dioxin-inducible), polypeptide 1; CYP1B1).

Cytochromes P450 are associated with inflammation and infection. Hepatic cytochrome P450 activities are profoundly affected by various infections and inflammatory stimuli, some of which are suppressed and some induced (Morgan, E.T. (1997) Drug Metab. Rev. 29:1129-1188). Effects observed in vivo can be mimicked by proinflammatory cytokines and interferons. Autoantibodies to two cytochrome P450 proteins were found in patients with autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), a polyglandular autoimmune syndrome (OMIM *240300 Autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy).

Mutations in cytochromes P450 have been linked to metabolic disorders, including congenital adrenal hyperplasia, the most common adrenal disorder of infancy and childhood; pseudovitamin D-deficiency rickets; cerebrotendinous xanthomatosis, a lipid storage disease characterized by progressive neurologic dysfunction, premature atherosclerosis, and cataracts; and an inherited resistance to the anticoagulant drugs coumarin and warfarin (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, Inc. New York, NY, pp. 1968-1970; Takeyama, K. et al. (1997) Science 277:1827-1830; Kitanaka, S. et al. (1998) N. Engl. J. Med. 338:653-661; OMIM *213700 Cerebrotendinous xanthomatosis; and OMIM #122700 Coumarin resistance). Extremely high levels of expression of the cytochrome P450 protein aromatase were found in a fibrolamellar hepatocellular carcinoma from a boy with severe gynecomastia (feminization) (Agarwal, V.R. (1998) J. Clin. Endocrinol. Metab. 83:1797-1800).

The cytochrome P450 catalytic cycle is completed through reduction of cytochrome P450 by NADPH cytochrome P450 reductase (CPR). Another microsomal electron transport system consisting of cytochrome b5 and NADPH cytochrome b5 reductase has been widely viewed as a minor contributor of electrons to the cytochrome P450 catalytic cycle. However, a recent report by Lamb, D.C. et al. (1999; FEBS Lett. 462:283-288) identifies a Candida albicans cytochrome P450 (CYP51) which can be efficiently reduced and supported by the microsomal cytochrome b5/NADPH cytochrome b5 reductase system. Therefore, there are likely many cytochromes P450 which are supported by this alternative electron donor system.

Cytochrome b5 reductase is also responsible for the reduction of oxidized hemoglobin (methemoglobin, or ferrihemoglobin, which is unable to carry oxygen) to the active hemoglobin (ferrohemoglobin) in red blood cells. Methemoglobinemia results when there is a high level of oxidant drugs or an abnormal hemoglobin (hemoglobin M) which is not efficiently reduced.

Methemoglobinemia can also result from a hereditary deficiency in red cell cytochrome b5 reductase (Reviewed in Mansour, A. and Lurie, A.A. (1993) Am. J. Hematol. 42:7-12).

Members of the cytochrome P450 family are also closely associated with vitamin D synthesis and catabolism. Vitamin D exists as two biologically equivalent prohormones, ergocalciferol (vitamin $D_2$), produced in plant tissues, and cholecalciferol (vitamin $D_3$), produced in animal tissues. The latter form, cholecalciferol, is formed upon the exposure of 7-dehydrocholesterol to near ultraviolet light (i.e., 290-310 nm), normally resulting from even minimal periods of skin exposure to sunlight (reviewed in Miller, W.L. and Portale, A.A. (2000) Trends Endocrinol. Metab. 11:315-319).

Both prohormone forms are further metabolized in the liver to 25-hydroxyvitamin D (25(OH)D) by the enzyme 25-hydroxylase. 25(OH)D is the most abundant precursor form of vitamin D which must be further metabolized in the kidney to the active form, 1α,25-dihydroxyvitamin D ($1\alpha,25(OH)_2D$), by the enzyme 25-hydroxyvitamin D 1α-hydroxylase (1α-hydroxylase). Regulation of $1\alpha,25(OH)_2D$ production is primarily at this final step in the synthetic pathway. The activity of 1α-hydroxylase depends upon several physiological factors including the circulating level of the enzyme product ($1\alpha,25(OH)_2D$) and the levels of parathyroid hormone (PTH), calcitonin, insulin, calcium, phosphorus, growth hormone, and prolactin. Furthermore, extrarenal 1α-hydroxylase activity has been reported, suggesting that tissue-specific, local regulation of 1α,25(OH)2D production may also be biologically important. The catalysis of $1\alpha,25(OH)_2D$ to 24,25-dihydroxyvitamin D ($24,25(OH)_2D$), involving the enzyme 25-hydroxyvitamin D 24-hydroxylase (24-hydroxylase), also occurs in the kidney. 24-hydroxylase can also use 25(OH)D as a substrate (Shinki, T. et al. (1997) Proc. Natl. Acad. Sci. U.S.A. 94:12920-12925; Miller, W.L. and Portale, A.A. supra; and references within).

Vitamin D 25-hydroxylase, 1α-hydroxylase, and 24-hydroxylase are all NADPH-dependent, type I (mitochondrial) cytochrome P450 enzymes that show a high degree of homology with other members of the family. Vitamin D 25-hydroxylase also shows a broad substrate specificity and may also perform 26-hydroxylation of bile acid intermediates and 25, 26, and 27-hydroxylation of cholesterol (Dilworth, F.J. et al. (1995) J. Biol. Chem. 270:16766-16774; Miller, W.L. and Portale, A.A. supra; and references within).

The active form of vitamin D ($1\alpha,25(OH)_2D$) is involved in calcium and phosphate homeostasis and promotes the differentiation of myeloid and skin cells. Vitamin D deficiency resulting from deficiencies in the enzymes involved in vitamin D metabolism (e.g., 1 α-hydroxylase) causes hypocalcemia, hypophosphatemia, and vitamin D-dependent (sensitive) rickets, a disease characterized by loss of bone density and distinctive clinical features, including bandy or bow leggedness accompanied by a waddling gait. Deficiencies in vitamin D 25-hydroxylase cause cerebrotendinous xanthomatosis, a lipid-storage disease characterized by the deposition of cholesterol

and cholestanol in the Achilles' tendons, brain, lungs, and many other tissues. The disease presents with progressive neurologic dysfunction, including postpubescent cerebellar ataxia, atherosclerosis, and cataracts. Vitamin D 25-hydroxylase deficiency does not result in rickets, suggesting the existence of alternative pathways for the synthesis of 25(OH)D (Griffin, J.E. and Zerwekh, J.E. (1983) J. Clin. Invest. 72:1190-1199; Gamblin, G.T. et al. (1985) J. Clin. Invest. 75:954-960; and Miller, W.L. and Portale, A.A. supra).

Ferredoxin and ferredoxin reductase are electron transport accessory proteins which support at least one human cytochrome P450 species, cytochrome P450c27 encoded by the CYP27 gene (Dilworth, F.J. et al. (1996) Biochem. J. 320:267-71). A Streptomyces griseus cytochrome P450, CYP104D1, was heterologously expressed in E. coli and found to be reduced by the endogenous ferredoxin and ferredoxin reductase enzymes (Taylor, M. et al. (1999) Biochem. Biophys. Res. Commun. 263:838-42), suggesting that many cytochrome P450 species may be supported by the ferredoxin/ferredoxin reductase pair. Ferredoxin reductase has also been found in a model drug metabolism system to reduce actinomycin D, an antitumor antibiotic, to a reactive free radical species (Flitter, W.D. and Mason, R.P. (1988) Arch. Biochem. Biophys. 267:632-639).

Flavin-containing monooxygenase (FMO):

Flavin-containing monooxygenases oxidize the nucleophilic nitrogen, sulfur, and phosphorus heteroatom of an exceptional range of substrates. Like cytochromes P450, FMOs are microsomal and use NADPH and $O_2$; there is also a great deal of substrate overlap with cytochromes P450. The tissue distribution of FMOs includes liver, kidney, and lung.

Isoforms of FMO in mammals include FMO1, FMO2, FMO3, FMO4, and FMO5, which are expressed in a tissue-specific manner. The isoforms differ in their substrate specificities and properties such as inhibition by various compounds and stereospecificity of reaction. FMOs have a 13 amino acid signature sequence, the components of which span the N-terminal two-thirds of the sequences and include the FAD binding region and the FATGY motif found in many N-hydroxylating enzymes (Stehr, M. et al. (1998) Trends Biochem. Sci. 23:56-57; PRINTS FMOXYGENASE Flavin-containing monooxygenase signature). Specific reactions include oxidation of nucleophilic tertiary amines to N-oxides, secondary amines to hydroxylamines and nitrones, primary amines to hydroxylamines and oximes, and sulfur-containing compounds and phosphines to S- and P-oxides. Hydrazines, iodides, selenides, and boron-containing compounds are also substrates. FMOs are more heat labile and less detergent-sensitive than cytochromes P450 in vitro though FMO isoforms vary in thermal stability and detergent sensitivity.

FMOs play important roles in the metabolism of several drugs and xenobiotics. FMO (FMO3 in liver) is predominantly responsible for metabolizing (S)-nicotine to (S)-nicotine N-1'-oxide, which is excreted in urine. FMO is also involved in S-oxygenation of cimetidine, an $H_2$-

antagonist widely used for the treatment of gastric ulcers. Liver-expressed forms of FMO are not under the same regulatory control as cytochrome P450. In rats, for example, phenobarbital treatment leads to the induction of cytochrome P450, but the repression of FMO1.

Lysyl oxidase

Lysyl oxidase (lysine 6-oxidase, LO) is a copper-dependent amine oxidase involved in the formation of connective tissue matrices by crosslinking collagen and elastin. LO is secreted as an N-glycosylated precursor protein of approximately 50 kDa and cleaved to the mature form of the enzyme by a metalloprotease, although the precursor form is also active. The copper atom in LO is involved in the transport of electrons to and from oxygen to facilitate the oxidative deamination of lysine residues in these extracellular matrix proteins. While the coordination of copper is essential to LO activity, insufficient dietary intake of copper does not influence the expression of the apoenzyme. However, the absence of the functional LO is linked to the skeletal and vascular tissue disorders that are associated with dietary copper deficiency. LO is also inhibited by a variety of semicarbazides, hydrazines, and amino nitrites, as well as heparin. Beta-aminopropionitrile is a commonly used inhibitor. LO activity is increased in response to ozone, cadmium, and elevated levels of hormones released in response to local tissue trauma, such as transforming growth factor-beta, platelet-derived growth factor, angiotensin II, and fibroblast growth factor. Abnormalities in LO activity have been linked to Menkes syndrome and occipital horn syndrome. Cytosolic forms of the enzyme have been implicated in abnormal cell proliferation (reviewed in Rucker, R.B. et al. (1998) Am. J. Clin. Nutr. 67:996S-1002S and Smith-Mungo, L.I. and Kagan, H.M. (1998) Matrix Biol. 16:387-398).

Dihydrofolate reductases

Dihydrofolate reductases (DHFR) are ubiquitous enzymes that catalyze the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate, an essential step in the de novo synthesis of glycine and purines as well as the conversion of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP). The basic reaction is as follows:

$$\text{7,8-dihydrofolate} + \text{NADPH} \rightarrow \text{5,6,7,8-tetrahydrofolate} + \text{NADP}^+$$

The enzymes can be inhibited by a number of dihydrofolate analogs, including trimethroprim and methotrexate. Since an abundance of dTMP is required for DNA synthesis, rapidly dividing cells require the activity of DHFR. The replication of DNA viruses (i.e., herpesvirus) also requires high levels of DHFR activity. As a result, drugs that target DHFR have been used for cancer chemotherapy and to inhibit DNA virus replication. (For similar reasons, thymidylate synthetases are also target enzymes.) Drugs that inhibit DHFR are preferentially cytotoxic for rapidly dividing cells (or DNA virus-infected cells) but have no specificity, resulting in the indiscriminate destruction of

dividing cells. Furthermore, cancer cells may become resistant to drugs such as methotrexate as a result of acquired transport defects or the duplication of one or more DHFR genes (Stryer, L. (1988) Biochemistry. W.H Freeman and Co., Inc. New York. pp. 511-5619).

Aldo/keto reductases

Aldo/keto reductases are monomeric NADPH-dependent oxidoreductases with broad substrate specificities (Bohren, K.M. et al. (1989) J. Biol. Chem. 264:9547-9551). These enzymes catalyze the reduction of carbonyl-containing compounds, including carbonyl-containing sugars and aromatic compounds, to the corresponding alcohols. Therefore, a variety of carbonyl-containing drugs and xenobiotics are likely metabolized by enzymes of this class.

One known reaction catalyzed by a family member, aldose reductase, is the reduction of glucose to sorbitol, which is then further metabolized to fructose by sorbitol dehydrogenase. Under normal conditions, the reduction of glucose to sorbitol is a minor pathway. In hyperglycemic states, however, the accumulation of sorbitol is implicated in the development of diabetic complications (OMIM *103880 Aldo-keto reductase family 1, member B1). Members of this enzyme family are also highly expressed in some liver cancers (Cao, D. et al. (1998) J. Biol. Chem. 273:11429-11435).

Alcohol dehydrogenases

Alcohol dehydrogenases (ADHs) oxidize simple alcohols to the corresponding aldehydes. ADH is a cytosolic enzyme, prefers the cofactor $NAD^+$, and also binds zinc ion. Liver contains the highest levels of ADH, with lower levels in kidney, lung, and the gastric mucosa.

Known ADH isoforms are dimeric proteins composed of 40 kDa subunits. There are five known gene loci which encode these subunits (a, b, g, p, c), and some of the loci have characterized allelic variants ($b_1$, $b_2$, $b_3$, $g_1$, $g_2$). The subunits can form homodimers and heterodimers; the subunit composition determines the specific properties of the active enzyme. The holoenzymes have therefore been categorized as Class I (subunit compositions aa, ab, ag, bg, gg), Class II (pp), and Class III (cc). Class I ADH isozymes oxidize ethanol and other small aliphatic alcohols, and are inhibited by pyrazole. Class II isozymes prefer longer chain aliphatic and aromatic alcohols, are unable to oxidize methanol, and are not inhibited by pyrazole. Class III isozymes prefer even longer chain aliphatic alcohols (five carbons and longer) and aromatic alcohols, and are not inhibited by pyrazole.

The short-chain alcohol dehydrogenases include a number of related enzymes with a variety of substrate specificities. Included in this group are the mammalian enzymes D-beta-hydroxybutyrate dehydrogenase, (R)-3-hydroxybutyrate dehydrogenase, 15-hydroxyprostaglandin dehydrogenase, NADPH-dependent carbonyl reductase, corticosteroid 11-beta-dehydrogenase, and estradiol 17-beta-dehydrogenase, as well as the bacterial enzymes acetoacetyl-CoA reductase, glucose 1-dehydrogenase, 3-beta-hydroxysteroid dehydrogenase, 20-beta-hydroxysteroid dehydrogenase, ribitol

dehydrogenase, 3-oxoacyl reductase, 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase, sorbitol-6-phosphate 2-dehydrogenase, 7-alpha-hydroxysteroid dehydrogenase, cis-1,2-dihydroxy-3,4-cyclohexadiene-1-carboxylate dehydrogenase, *cis*-toluene dihydrodiol dehydrogenase, *cis*-benzene glycol dehydrogenase, biphenyl-2,3-dihydro-2,3-diol dehydrogenase, N-acylmannosamine 1-dehydrogenase, and 2-deoxy-D-gluconate 3-dehydrogenase (Krozowski, Z. (1994) J. Steroid Biochem. Mol. Biol. 51:125-130; Krozowski, Z. (1992) Mol. Cell Endocrinol. 84:C25-31; and Marks, A.R. et al. (1992) J. Biol. Chem. 267:15459-15463).

Sulfotransferase

Sulfate conjugation occurs on many of the same substrates which undergo O-glucuronidation to produce a highly water-soluble sulfuric acid ester. Sulfotransferases (ST) catalyze this reaction by transferring $SO_3^-$ from the cofactor 3'-phosphoadenosine-5'-phosphosulfate (PAPS) to the substrate. ST substrates are predominantly phenols and aliphatic alcohols, but also include aromatic amines and aliphatic amines, which are conjugated to produce the corresponding sulfamates. The products of these reactions are excreted mainly in urine.

STs are found in a wide range of tissues, including liver, kidney, intestinal tract, lung, platelets, and brain. The enzymes are generally cytosolic, and multiple forms are often co-expressed. For example, there are more than a dozen forms of ST in rat liver cytosol. These biochemically characterized STs fall into five classes based on their substrate preference: arylsulfotransferase, alcohol sulfotransferase, estrogen sulfotransferase, tyrosine ester sulfotransferase, and bile salt sulfotransferase.

ST enzyme activity varies greatly with sex and age in rats. The combined effects of developmental cues and sex-related hormones are thought to lead to these differences in ST expression profiles, as well as the profiles of other DMEs such as cytochromes P450. Notably, the high expression of STs in cats partially compensates for their low level of UDP glucuronyltransferase activity.

Several forms of ST have been purified from human liver cytosol and cloned. There are two phenol sulfotransferases with different thermal stabilities and substrate preferences. The thermostable enzyme catalyzes the sulfation of phenols such as para-nitrophenol, minoxidil, and acetaminophen; the thermolabile enzyme prefers monoamine substrates such as dopamine, epinephrine, and levadopa. Other cloned STs include an estrogen sulfotransferase and an N-acetylglucosamine-6-O-sulfotransferase. This last enzyme is illustrative of the other major role of STs in cellular biochemistry, the modification of carbohydrate structures that may be important in cellular differentiation and maturation of proteoglycans. Indeed, an inherited defect in a sulfotransferase has been implicated in macular corneal dystrophy, a disorder characterized by a failure to synthesize mature keratan sulfate proteoglycans (Nakazawa, K. et al. (1984) J. Biol. Chem.

259:13751-13757; OMIM *217800 Macular dystrophy, corneal).

Galactosyltransferases

Galactosyltransferases are a subset of glycosyltransferases that transfer galactose (Gal) to the terminal N-acetylglucosamine (GlcNAc) oligosaccharide chains that are part of glycoproteins or glycolipids that are free in solution (Kolbinger, F. et al. (1998) J. Biol. Chem. 273:433-440; Amado, M. et al. (1999) Biochim. Biophys. Acta 1473:35-53). Galactosyltransferases have been detected on the cell surface and as soluble extracellular proteins, in addition to being present in the Golgi. β1,3-galactosyltransferases form Type I carbohydrate chains with Gal (β1-3)GlcNAc linkages. Known human and mouse β1,3-galactosyltransferases appear to have a short cytosolic domain, a single transmembrane domain, and a catalytic domain with eight conserved regions. (Kolbinger, supra and Hennet, T. et al. (1998) J. Biol. Chem. 273:58-65). In mouse UDP-galactose:β-N-acetylglucosamine β1,3-galactosyltransferase-I region 1 is located at amino acid residues 78-83, region 2 is located at amino acid residues 93-102, region 3 is located at amino acid residues 116-119, region 4 is located at amino acid residues 147-158, region 5 is located at amino acid residues 172-183, region 6 is located at amino acid residues 203-206, region 7 is located at amino acid residues 236-246, and region 8 is located at amino acid residues 264-275. A variant of a sequence found within mouse UDP-galactose:β-N-acetylglucosamine β1,3-galactosyltransferase-I region 8 is also found in bacterial galactosyltransferases, suggesting that this sequence defines a galactosyltransferase sequence motif (Hennet, supra). Recent work suggests that brainiac protein is a β1,3-galactosyltransferase (Yuan, Y. et al. (1997) Cell 88:9-11; and Hennet, supra).

UDP-Gal:GlcNAc-1,4-galactosyltransferase (-1,4-GalT) (Sato, T. et al., (1997) EMBO J. 16:1850-1857) catalyzes the formation of Type II carbohydrate chains with Gal (β1-4)GlcNAc linkages. As is the case with the β1,3-galactosyltransferase, a soluble form of the enzyme is formed by cleavage of the membrane-bound form. Amino acids conserved among β1,4-galactosyltransferases include two cysteines linked through a disulfide-bond and a putative UDP-galactose-binding site in the catalytic domain (Yadav, S. and Brew, K. (1990) J. Biol. Chem. 265:14163-14169; Yadav, S.P. and Brew, K. (1991) J. Biol. Chem. 266:698-703; and Shaper, N.L. et al. (1997) J. Biol. Chem. 272:31389-31399). β1,4-galactosyltransferases have several specialized roles in addition to synthesizing carbohydrate chains on glycoproteins or glycolipids. In mammals a β1,4-galactosyltransferase, as part of a heterodimer with α-lactalbumin, functions in lactating mammary gland lactose production. A β1,4-galactosyltransferase on the surface of sperm functions as a receptor that specifically recognizes the egg. Cell surface β1,4-galactosyltransferases also function in cell adhesion, cell/basal lamina interaction, and normal and metastatic cell migration. (Shur, B. (1993) Curr. Opin. Cell Biol. 5:854-863; and Shaper, J. (1995) Adv. Exp. Med. Biol. 376:95-104).

Gamma-glutamyl transpeptidase

Gamma-glutamyl transpeptidases are ubiquitously expressed enzymes that initiate extracellular glutathione (GSH) breakdown by cleaving gamma-glutamyl amide bonds. The breakdown of GSH provides cells with a regional cysteine pool for biosynthetic pathways. Gamma-glutamyl transpeptidases also contribute to cellular antioxidant defenses and expression is induced by oxidative stress. The cell surface-localized glycoproteins.are expressed at high levels in cancer cells. Studies have suggested that the high level of gamma-glutamyl transpeptidase activity present on the surface of cancer cells could be exploited to activate precursor drugs, resulting in high local concentrations of anti-cancer therapeutic agents (Hanigan, M.H. (1998) Chem. Biol. Interact. 111-112:333-42; Taniguchi, N. and Ikeda, Y. (1998) Adv. Enzymol. Relat. Areas Mol. Biol. 72:239-78; Chikhi, N. et al. (1999) Comp. Biochem. Physiol. B. Biochem. Mol. Biol. 122:367-380).

Aminotransferases

Aminotransferases comprise a family of pyridoxal 5'-phosphate (PLP) -dependent enzymes that catalyze transformations of amino acids. Aspartate aminotransferase (AspAT) is the most extensively studied PLP-containing enzyme. It catalyzes the reversible transamination of dicarboxylic L-amino acids, aspartate and glutamate, and the corresponding 2-oxo acids, oxalacetate and 2-oxoglutarate. Other members of the family include pyruvate aminotransferase, branched-chain amino acid aminotransferase, tyrosine aminotransferase, aromatic aminotransferase, alanine:glyoxylate aminotransferase (AGT), and kynurenine aminotransferase (Vacca, R.A. et al. (1997) J. Biol. Chem. 272:21932-21937).

Primary hyperoxaluria type-1 is an autosomal recessive disorder resulting in a deficiency in the liver-specific peroxisomal enzyme, alanine:glyoxylate aminotransferase-1. The phenotype of the disorder is a deficiency in glyoxylate metabolism. In the absence of AGT, glyoxylate is oxidized to oxalate rather than being transaminated to glycine. The result is the deposition of insoluble calcium oxalate in the kidneys and urinary tract, ultimately causing renal failure (Lumb, M.J. et al. (1999) J. Biol. Chem. 274:20587-20596).

Kynurenine aminotransferase catalyzes the irreversible transamination of the L-tryptophan metabolite L-kynurenine to form kynurenic acid. The enzyme may also catalyze the reversible transamination reaction between L-2-aminoadipate and 2-oxoglutarate to produce 2-oxoadipate and L-glutamate. Kynurenic acid is a putative modulator of glutamatergic neurotransmission; thus a deficiency in kynurenine aminotransferase may be associated with pleotrophic effects (Buchli, R. et al. (1995) J. Biol. Chem. 270:29330-29335).

Catechol-*O*-methyltransferase

Catechol-*O*-methyltransferase (COMT) catalyzes the transfer of the methyl group of *S*-adenosyl-L-methionine (AdoMet; SAM) donor to one of the hydroxyl groups of the catechol

substrate (e.g., L-dopa, dopamine, or DBA). Methylation of the 3'-hydroxyl group is favored over methylation of the 4'-hydroxyl group and the membrane bound isoform of COMT is more regiospecific than the soluble form. Translation of the soluble form of the enzyme results from utilization of an internal start codon in a full-length mRNA (1.5 kb) or from the translation of a shorter mRNA (1.3 kb), transcribed from an internal promoter. The proposed $S_N2$-like methylation reaction requires $Mg^{++}$ and is inhibited by $Ca^{++}$. The binding of the donor and substrate to COMT occurs sequentially. AdoMet first binds COMT in a $Mg^{++}$-independent manner, followed by the binding of $Mg^{++}$ and the binding of the catechol substrate.

The amount of COMT in tissues is relatively high compared to the amount of activity normally required, thus inhibition is problematic. Nonetheless, inhibitors have been developed for in vitro use (e.g., gallates, tropolone, U-0521, and 3',4'-dihydroxy-2-methyl-propiophetropolone) and for clinical use (e.g., nitrocatechol-based compounds and tolcapone). Administration of these inhibitors results in the increased half-life of L-dopa and the consequent formation of dopamine. Inhibition of COMT is also likely to increase the half-life of various other catechol-structure compounds, including but not limited to epinephrine/norepinephrine, isoprenaline, rimiterol, dobutamine, fenoldopam, apomorphine, and α-methyldopa. A deficiency in norepinephrine has been linked to clinical depression, hence the use of COMT inhibitors could be usefull in the treatment of depression. COMT inhibitors are generally well tolerated with minimal side effects and are ultimately metabolized in the liver with only minor accumulation of metabolites in the body (Männistö, P.T. and Kaakkola, S. (1999) Pharmacol. Rev. 51:593-628).

Copper-zinc superoxide dismutases

Copper-zinc superoxide dismutases are compact homodimeric metalloenzymes involved in cellular defenses against oxidative damage. The enzymes contain one atom of zinc and one atom of copper per subunit and catalyze the dismutation of superoxide anions into $O_2$ and $H_2O_2$. The rate of dismutation is diffusion-limited and consequently enhanced by the presence of favorable electrostatic interactions between the substrate and enzyme active site. Examples of this class of enzyme have been identified in the cytoplasm of all the eukaryotic cells as well as in the periplasm of several bacterial species. Copper-zinc superoxide dismutases are robust enzymes that are highly resistant to proteolytic digestion and denaturing by urea and SDS. In addition to the compact structure of the enzymes, the presence of the metal ions and intrasubunit disulfide bonds is believed to be responsible for enzyme stability. The enzymes undergo reversible denaturation at temperatures as high as 70°C (Battistoni, A. et al. (1998) J. Biol. Chem. 273:5655-5661).

Overexpression of superoxide dismutase has been implicated in enhancing freezing tolerance of transgenic alfalfa as well as providing resistance to environmental toxins such as the diphenyl ether herbicide, acifluorfen (McKersie, B.D. et al. (1993) Plant Physiol. 103:1155-1163). In addtion,

yeast cells become more resistant to freeze-thaw damage following exposure to hydrogen peroxide which causes the yeast cells to adapt to further peroxide stress by upregulating expression of superoxide dismutases. In this study, mutations to yeast superoxide dismutase genes had a more detrimental effect on freeze-thaw resistance than mutations which affected the regulation of glutathione metabolism, long suspected of being important in determining an organism's survival through the process of cryopreservation (Jong-In Park, J.-I. et al. (1998) J. Biol. Chem. 273:22921-22928).

Expression of superoxide dismutase is also associated with Mycobacterium tuberculosis, the organism that causes tuberculosis. Superoxide dismutase is one of the ten major proteins secreted by M. tuberculosis and its expression is upregulated approximately 5-fold in response to oxidative stress. M. tuberculosis expresses almost two orders of magnitude more superoxide dismutase than the nonpathogenic mycobacterium M. smegmatis, and secretes a much higher proportion of the expressed enzyme. The result is the secretion of ~350-fold more enzyme by M. tuberculosis than M. smegmatis, providing substantial resistance to oxidative stress (Harth, G. and Horwitz, M.A. (1999) J. Biol. Chem. 274:4281-4292).

The reduced expression of copper-zinc superoxide dismutases, as well as other enzymes with anti-oxidant capabilities, has been implicated in the early stages of cancer. The expression of copper-zinc superoxide dismutases is reduced in prostatic intraepithelial neoplasia and prostate carcinomas, (Bostwick, D.G. (2000) Cancer 89:123-134).

Phosphoesterases

Phosphotriesterases (PTE, paraoxonases) are enzymes that hydrolyze toxic organophosphorus compounds and have been isolated from a variety of tissues. Phosphotriesterases play a central role in the detoxification of insecticides by mammals. Birds and insects lack PTE, and as a result have reduced tolerance for organophosphorus compounds (Vilanova, E. and Sogorb, M.A. (1999) Crit. Rev. Toxicol. 29:21-57). Phosphotriesterase activity varies among individuals and is lower in infants than adults. PTE knockout mice are markedly more sensitive to the organophosphate-based toxins diazoxon and chlorpyrifos oxon (Furlong, C.E., et al. (2000) Neurotoxicology 21:91-100). Phosphotriesterase is also implicated in atherosclerosis and diseases involving lipoprotein metabolism.

Glycerophosphoryl diester phosphodiesterase (also known as glycerophosphodiester phosphodiesterase) is a phosphodiesterase which hydrolyzes deacetylated phospholipid glycerophosphodiesters to produce sn-glycerol-3-phosphate and an alcohol. Glycerophosphocholine, glycerophosphoethanolamine, glycerophosphoglycerol, and glycerophosphoinositol are examples of substrates for glycerophosphoryl diester phosphodiesterases. A glycerophosphoryl diester phosphodiesterase from E. coli has broad specificity for glycerophosphodiester substrates (Larson,

T.J. et al. (1983) J. Biol. Chem. 248:5428-5432).

Cyclic nucleotide phosphodiesterases (PDEs) are crucial enzymes in the regulation of the cyclic nucleotides cAMP and cGMP. cAMP and cGMP function as intracellular second messengers to transduce a variety of extracellular signals including hormones, light, and neurotransmitters. PDEs degrade cyclic nucleotides to their corresponding monophosphates, thereby regulating the intracellular concentrations of cyclic nucleotides and their effects on signal transduction. Due to their roles as regulators of signal transduction, PDEs have been extensively studied as chemotherapeutic targets (Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481; Torphy, J.T. (1998) Am. J. Resp. Crit. Care Med. 157:351-370).

Families of mammalian PDEs have been classified based on their substrate specificity and affinity, sensitivity to cofactors, and sensitivity to inhibitory agents (Beavo, J.A. (1995) Physiol. Rev. 75:725-748; Conti, M. et al. (1995) Endocrine Rev. 16:370-389). Several of these families contain distinct genes, many of which are expressed in different tissues as splice variants. Within PDE families, there are multiple isozymes and multiple splice variants of these isozymes (Conti, M. and S.-L.C. Jin (1999) Prog. Nucleic Acid Res. Mol. Biol. 63:1-38). The existence of multiple PDE families, isozymes, and splice variants is an indication of the variety and complexity of the regulatory pathways involving cyclic nucleotides (Houslay, M.D. and G. Milligan (1997) Trends Biochem. Sci. 22:217-224).

Type 1 PDEs (PDE1s) are $Ca^{2+}$/calmodulin-dependent and appear to be encoded by at least three different genes, each having at least two different splice variants (Kakkar, R. et al. (1999) Cell Mol. Life Sci. 55:1164-1186). PDE1s have been found in the lung, heart, and brain. Some PDE1 isozymes are regulated in vitro by phosphorylation/dephosphorylation. Phosphorylation of these PDE1 isozymes decreases the affinity of the enzyme for calmodulin, decreases PDE activity, and increases steady state levels of cAMP (Kakkar, supra). PDE1s may provide useful therapeutic targets for disorders of the central nervous system and the cardiovascular and immune systems, due to the involvement of PDE1s in both cyclic nucleotide and calcium signaling (Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481).

PDE2s are cGMP-stimulated PDEs that have been found in the cerebellum, neocortex, heart, kidney, lung, pulmonary artery, and skeletal muscle (Sadhu, K. et al. (1999) J. Histochem. Cytochem. 47:895-906). PDE2s are thought to mediate the effects of cAMP on catecholamine secretion, participate in the regulation of aldosterone (Beavo, supra), and play a role in olfactory signal transduction (Juilfs, D.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:3388-3395).

PDE3s have high affinity for both cGMP and cAMP, and so these cyclic nucleotides act as competitive substrates for PDE3s. PDE3s play roles in stimulating myocardial contractility, inhibiting platelet aggregation, relaxing vascular and airway smooth muscle, inhibiting proliferation

of T-lymphocytes and cultured vascular smooth muscle cells, and regulating catecholamine-induced release of free fatty acids from adipose tissue. The PDE3 family of phosphodiesterases are sensitive to specific inhibitors such as cilostamide, enoximone, and lixazinone. Isozymes of PDE3 can be regulated by cAMP-dependent protein kinase, or by insulin-dependent kinases (Degerman, E. et al. (1997) J. Biol. Chem. 272:6823-6826).

PDE4s are specific for cAMP; are localized to airway smooth muscle, the vascular endothelium, and all inflammatory cells; and can be activated by cAMP-dependent phosphorylation. Since elevation of cAMP levels can lead to suppression of inflammatory cell activation and to relaxation of bronchial smooth muscle, PDE4s have been studied extensively as possible targets for novel anti-inflammatory agents, with special emphasis placed on the discovery of asthma treatments. PDE4 inhibitors are currently undergoing clinical trials as treatments for asthma, chronic obstructive pulmonary disease, and atopic eczema. All four known isozymes of PDE4 are susceptible to the inhibitor rolipram, a compound which has been shown to improve behavioral memory in mice (Barad, M. et al. (1998) Proc. Natl. Acad. Sci. USA 95:15020-15025). PDE4 inhibitors have also been studied as possible therapeutic agents against acute lung injury, endotoxemia, rheumatoid arthritis, multiple sclerosis, and various neurological and gastrointestinal indications (Doherty, A.M. (1999) Curr. Opin. Chem. Biol. 3:466-473).

PDE5 is highly selective for cGMP as a substrate (Turko, I.V. et al. (1998) Biochemistry 37:4200-4205), and has two allosteric cGMP-specific binding sites (McAllister-Lucas, L.M. et al. (1995) J. Biol. Chem. 270:30671-30679). Binding of cGMP to these allosteric binding sites seems to be important for phosphorylation of PDE5 by cGMP-dependent protein kinase rather than for direct regulation of catalytic activity. High levels of PDE5 are found in vascular smooth muscle, platelets, lung, and kidney. The inhibitor zaprinast is effective against PDE5 and PDE1s. Modification of zaprinast to provide specificity against PDE5 has resulted in sildenafil (VIAGRA; Pfizer, Inc., New York NY), a treatment for male erectile dysfunction (Terrett, N. et al. (1996) Bioorg. Med. Chem. Lett. 6:1819-1824). Inhibitors of PDE5 are currently being studied as agents for cardiovascular therapy (Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481).

PDE6s, the photoreceptor cyclic nucleotide phosphodiesterases, are crucial components of the phototransduction cascade. In association with the G-protein transducin, PDE6s hydrolyze cGMP to regulate cGMP-gated cation channels in photoreceptor membranes. In addition to the cGMP-binding active site, PDE6s also have two high-affinity cGMP-binding sites which are thought to play a regulatory role in PDE6 function (Artemyev, N.O. et al. (1998) Methods 14:93-104). Defects in PDE6s have been associated with retinal disease. Retinal degeneration in the rd mouse (Yan, W. et al. (1998) Invest. Opthalmol. Vis. Sci. 39:2529-2536), autosomal recessive retinitis pigmentosa in humans (Danciger, M. et al. (1995) Genomics 30:1-12), and rod/cone dysplasia 1 in Irish Setter dogs

(Suber, M.L. et al. (1993) Proc. Natl. Acad. Sci. USA 90:3968-3972) have been attributed to mutations in the PDE6B gene.

The PDE7 family of PDEs consists of only one known member having multiple splice variants (Bloom, T.J. and J.A. Beavo (1996) Proc. Natl. Acad. Sci. USA 93:141813-24192). PDE7s are cAMP specific, but little else is known about their physiological function. Although mRNAs encoding PDE7s are found in skeletal muscle, heart, brain, lung, kidney, and pancreas, expression of PDE7 proteins is restricted to specific tissue types (Han, P. et al. (1997) J. Biol. Chem. 272:16152-16157; Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481). PDE7s are very closely related to the PDE4 family; however, PDE7s are not inhibited by rolipram, a specific inhibitor of PDE4s (Beavo, supra).

PDE8s are cAMP specific, and are closely related to the PDE4 family. PDE8s are expressed in thyroid gland, testis, eye, liver, skeletal muscle, heart, kidney, ovary, and brain. The cAMP-hydrolyzing activity of PDE8s is not inhibited by the PDE inhibitors rolipram, vinpocetine, milrinone, IBMX (3-isobutyl-1-methylxanthine), or zaprinast, but PDE8s are inhibited by dipyridamole (Fisher, D.A. et al. (1998) Biochem. Biophys. Res. Commun. 246:570-577; Hayashi, M. et al. (1998) Biochem. Biophys. Res. Commun. 250:751-1256; Soderling, S.H. et al. (1998) Proc. Natl. Acad. Sci. USA 95:8991-8996).

PDE9s are cGMP specific and most closely resemble the PDE8 family of PDEs. PDE9s are expressed in kidney, liver, lung, brain, spleen, and small intestine. PDE9s are not inhibited by sildenafil (VIAGRA; Pfizer, Inc., New York NY), rolipram, vinpocetine, dipyridamole, or IBMX (3-isobutyl-1-methylxanthine), but they are sensitive to the PDE5 inhibitor zaprinast (Fisher, D.A. et al. (1998) J. Biol. Chem. 273:15559-15564; Soderling, S.H. et al. (1998) J. Biol. Chem. 273:15553-15558).

PDE10s are dual-substrate PDEs, hydrolyzing both cAMP and cGMP. PDE10s are expressed in brain, thyroid, and testis. (Soderling, S.H. et al. (1999) Proc. Natl. Acad. Sci. USA 96:7071-12076; Fujishige, K. et al. (1999) J. Biol. Chem. 274:18438-18445; Loughney, K. et al (1999) Gene 234:109-117).

PDEs are composed of a catalytic domain of about 270-300 amino acids, an N-terminal regulatory domain responsible for binding cofactors, and, in some cases, a hydrophilic C-terminal domain of unknown function (Conti, M. and S.-L.C. Jin (1999) Prog. Nucleic Acid Res. Mol. Biol. 63:1-38). A conserved, putative zinc-binding motif has been identified in the catalytic domain of all PDEs. N-terminal regulatory domains include non-catalytic cGMP-binding domains in PDE2s, PDE5s, and PDE6s; calmodulin-binding domains in PDE1s; and domains containing phosphorylation sites in PDE3s and PDE4s. In PDE5, the N-terminal cGMP-binding domain spans about 380 amino acid residues and comprises tandem repeats of a conserved sequence motif (McAllister-Lucas, L.M.

et al. (1993) J. Biol. Chem. 268:22863-22873). The NKXnD motif has been shown by mutagenesis to be important for cGMP binding (Turko, I.V. et al. (1996) J. Biol. Chem. 271:22240-22244). PDE families display approximately 30% amino acid identity within the catalytic domain; however, isozymes within the same family typically display about 85-95% identity in this region (e.g. PDE4A vs PDE4B). Furthermore, within a family there is extensive similarity (>60%) outside the catalytic domain; while across families, there is little or no sequence similarity outside this domain.

Many of the constituent functions of immune and inflammatory responses are inhibited by agents that increase intracellular levels of cAMP (Verghese, M.W. et al. (1995) Mol. Pharmacol. 47:1164-1171). A variety of diseases have been attributed to increased PDE activity and associated with decreased levels of cyclic nucleotides. For example, a form of diabetes insipidus in mice has been associated with increased PDE4 activity, an increase in low-$K_m$ cAMP PDE activity has been reported in leukocytes of atopic patients, and PDE3 has been associated with cardiac disease.

Many inhibitors of PDEs have undergone clinical evaluation (Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481; Torphy, T.J. (1998) Am. J. Respir. Crit. Care Med. 157:351-370). PDE3 inhibitors are being developed as antithrombotic agents, antihypertensive agents, and as cardiotonic agents useful in the treatment of congestive heart failure. Rolipram, a PDE4 inhibitor, has been used in the treatment of depression, and other PDE4 inhibitors have an anti-inflammatory effect. Rolipram may inhibit HIV-1 replication (Angel, J.B. et al. (1995) AIDS 9:1137-1144). Additionally, rolipram suppresses the production of cytokines such as TNF-$\alpha$ and $\beta$ and interferon $\gamma$, and thus is effective against encephalomyelitis. Rolipram may also be effective in treating tardive dyskinesia and multiple sclerosis (Sommer, N. et al. (1995) Nat. Med. 1:244-248; Sasaki, H. et al. (1995) Eur. J. Pharmacol. 282:71-76). Theophylline is a nonspecific PDE inhibitor used in treatment of bronchial asthma and other respiratory diseases. Theophylline is believed to act on airway smooth muscle function and in an anti-inflammatory or immunomodulatory capacity (Banner, K.H. and C.P. Page (1995) Eur. Respir. J. 8:996-1000). Pentoxifylline is another nonspecific PDE inhibitor used in the treatment of intermittent claudication and diabetes-induced peripheral vascular disease. Pentoxifylline is also known to block TNF-$\alpha$ production and may inhibit HIV-1 replication (Angel et al., supra).

PDEs have been reported to affect cellular proliferation of a variety of cell types (Conti et al. (1995) Endocrine Rev. 16:370-389) and have been implicated in various cancers. Growth of prostate carcinoma cell lines DU145 and LNCaP was inhibited by delivery of cAMP derivatives and PDE inhibitors (Bang, Y.J. et al. (1994) Proc. Natl. Acad. Sci. USA 91:5330-5334). These cells also showed a permanent conversion in phenotype from epithelial to neuronal morphology. It has also been suggested that PDE inhibitors can regulate mesangial cell proliferation (Matousovic, K. et al. (1995) J. Clin. Invest. 96:401-410) and lymphocyte proliferation (Joulain, C. et al. (1995) J. Lipid

Mediat. Cell Signal. 11:63-79). One cancer treatment involves intracellular delivery of PDEs to particular cellular compartments of tumors, resulting in cell death (Deonarain, M.P. and A.A. Epenetos (1994) Br. J. Cancer 70:786-794).

Members of the UDP glucuronyltransferase family (UGTs) catalyze the transfer of a glucuronic acid group from the cofactor uridine diphosphate-glucuronic acid (UDP-glucuronic acid) to a substrate. The transfer is generally to a nucleophilic heteroatom (O, N, or S). Substrates include xenobiotics which have been functionalized by Phase I reactions, as well as endogenous compounds such as bilirubin, steroid hormones, and thyroid hormones. Products of glucuronidation are excreted in urine if the molecular weight of the substrate is less than about 250 g/mol, whereas larger glucuronidated substrates are excreted in bile.

UGTs are located in the microsomes of liver, kidney, intestine, skin, brain, spleen, and nasal mucosa, where they are on the same side of the endoplasmic reticulum membrane as cytochrome P450 enzymes and flavin-containing monooxygenases. UGTs have a C-terminal membrane-spanning domain which anchors them in the endoplasmic reticulum membrane, and a conserved signature domain of about 50 amino acid residues in their C terminal section (PROSITE PDOC00359 UDP-glycosyltransferase signature).

UGTs involved in drug metabolism are encoded by two gene families, UGT1 and UGT2. Members of the UGT1 family result from alternative splicing of a single gene locus, which has a variable substrate binding domain and constant region involved in cofactor binding and membrane insertion. Members of the UGT2 family are encoded by separate gene loci, and are divided into two families, UGT2A and UGT2B. The 2A subfamily is expressed in olfactory epithelium, and the 2B subfamily is expressed in liver microsomes. Mutations in UGT genes are associated with hyperbilirubinemia (OMIM #143500 Hyperbilirubinemia I); Crigler-Najjar syndrome, characterized by intense hyperbilirubinemia from birth (OMIM #218800 Crigler-Najjar syndrome); and a milder form of hyperbilirubinemia termed Gilbert's disease (OMIM *191740 UGT1).

Thioesterases

Two soluble thioesterases involved in fatty acid biosynthesis have been isolated from mammalian tissues, one which is active only toward long-chain fatty-acyl thioesters and one which is active toward thioesters with a wide range of fatty-acyl chain-lengths. These thioesterases catalyze the chain-terminating step in the *de novo* biosynthesis of fatty acids. Chain termination involves the hydrolysis of the thioester bond which links the fatty acyl chain to the 4'-phosphopantetheine prosthetic group of the acyl carrier protein (ACP) subunit of the fatty acid synthase (Smith, S. (1981a) Methods Enzymol. 71:181-188; Smith, S. (1981b) Methods Enzymol. 71:188-200).

E. coli contains two soluble thioesterases, thioesterase I which is active only toward long-chain acyl thioesters, and thioesterase II (TEII) which has a broad chain-length specificity (Naggert,

J. et al. (1991) J. Biol. Chem. 266:11044-11050). E. coli TEII does not exhibit sequence similarity with either of the two types of mammalian thioesterases which function as chain-terminating enzymes in *de novo* fatty acid biosynthesis. Unlike the mammalian thioesterases, E. coli TEII lacks the characteristic serine active site gly-X-ser-X-gly sequence motif and is not inactivated by the serine modifying agent diisopropyl fluorophosphate. However, modification of histidine 58 by iodoacetamide and diethylpyrocarbonate abolished TEII activity. Overexpression of TEII did not alter fatty acid content in E. coli, which suggests that it does not function as a chain-terminating enzyme in fatty acid biosynthesis (Naggert et al., supra). For that reason, Naggert et al. (supra) proposed that the physiological substrates for E. coli TEII may be coenzyme A (CoA)-fatty acid esters instead of ACP-phosphopanthetheine-fatty acid esters.

Carboxylesterases

Mammalian carboxylesterases are a multigene family expressed in a variety of tissues and cell types. Acetylcholinesterase, butyrylcholinesterase, and carboxylesterase are grouped into the serine superfamily of esterases (B-esterases). Other carboxylesterases include thyroglobulin, thrombin, Factor IX, gliotactin, and plasminogen. Carboxylesterases catalyze the hydrolysis of ester- and amide- groups from molecules and are involved in detoxification of drugs, environmental toxins, and carcinogens. Substrates for carboxylesterases include short- and long-chain acyl-glycerols, acylcarnitine, carbonates, dipivefrin hydrochloride, cocaine, salicylates, capsaicin, palmitoyl-coenzyme A, imidapril, haloperidol, pyrrolizidine alkaloids, steroids, p-nitrophenyl acetate, malathion, butanilicaine, and isocarboxazide. Carboxylesterases are also important for the conversion of prodrugs to free acids, which may be the active form of the drug (e.g., lovastatin, used to lower blood cholesterol) (reviewed in Satoh, T. and Hosokawa, M. (1998) Annu. Rev. Pharmacol. Toxicol.38:257-288). Neuroligins are a class of molecules that (i) have N-terminal signal sequences, (ii) resemble cell-surface receptors, (iii) contain carboxylesterase domains, (iv) are highly expressed in the brain, and (v) bind to neurexins in a calcium-dependent manner. Despite the homology to carboxylesterases, neuroligins lack the active site serine residue, implying a role in substrate binding rather than catalysis (Ichtchenko, K. et al. (1996) J. Biol. Chem. 271:2676-2682).

Squalene epoxidase

Squalene epoxidase (squalene monooxygenase, SE) is a microsomal membrane-bound, FAD-dependent oxidoreductase that catalyzes the first oxygenation step in the sterol biosynthetic pathway of eukaryotic cells. Cholesterol is an essential structural component of cytoplasmic membranes acquired via the LDL receptor-mediated pathway or the biosynthetic pathway. SE converts squalene to 2,3(*S*)-oxidosqualene, which is then converted to lanosterol and then cholesterol.

High serum cholesterol levels result in the formation of atherosclerotic plaques in the arteries of higher organisms. This deposition of highly insoluble lipid material onto the walls of essential

blood vessels results in decreased blood flow and potential necrosis. HMG-CoA reductase is responsible for the first committed step in cholesterol biosynthesis, conversion of 3-hydroxyl-3-methyl-glutaryl CoA (HMG-CoA) to mevalonate. HMG-CoA is the target of a number of pharmaceutical compounds designed to lower plasma cholesterol levels, but inhibition of MHG-CoA also results in the reduced synthesis of non-sterol intermediates required for other biochemical pathways. Since SE catalyzes a rate-limiting reaction that occurs later in the sterol synthesis pathway with cholesterol as the only end product, SE is a better ideal target for the design of anti-hyperlipidemic drugs (Nakamura, Y. et al. (1996) 271:8053-8056).

Epoxide hydrolases

Epoxide hydrolases catalyze the addition of water to epoxide-containing compounds, thereby hydrolyzing epoxides to their corresponding 1,2-diols. They are related to bacterial haloalkane dehalogenases and show sequence similarity to other members of the $\alpha/\beta$ hydrolase fold family of enzymes. This family of enzymes is important for the detoxification of xenobiotic epoxide compounds which are often highly electrophilic and destructive when introduced. Examples of epoxide hydrolase reactions include the hydrolysis of some leukotoxin to leukotoxin diol, and isoleukotoxin to isoleukotoxin diol. Leukotoxins alter membrane permeability and ion transport and cause inflammatory responses. In addition, epoxide carcinogens are produced by cytochrome P450 as intermediates in the detoxification of drugs and environmental toxins. Epoxide hydrolases possess a catalytic triad composed of Asp, Asp, and His (Arand, M. et al. (1996) J. Biol. Chem. 271:4223-4229; Rink, R. et al. (1997) J. Biol. Chem. 272:14650-14657; Argiriadi, M.A. et al. (2000) J. Biol. Chem. 275:15265-15270).

Enzymes involved in tyrosine catalysis

The degradation of the amino acid tyrosine, to either succinate and pyruvate or fumarate and acetoacetate, requires a large number of enzymes and generates a large number of intermediate compounds. In addition, many xenobiotic compounds may be metabolized using one or more reactions that are part of the tyrosine catabolic pathway. Enzymes involved in the degradation of tyrosine to succinate and pyruvate (e.g., in *Arthrobacter* species) include 4-hydroxyphenylpyruvate oxidase, 4-hydroxyphenylacetate 3-hydroxylase, 3,4-dihydroxyphenylacetate 2,3-dioxygenase, 5-carboxymethyl-2-hydroxymuconic semialdehyde dehydrogenase, *trans,cis*-5-carboxymethyl-2-hydroxymuconate isomerase, homoprotocatechuate isomerase/decarboxylase, *cis*-2-oxohept-3-ene-1,7-dioate hydratase, 2,4-dihydroxyhept-*trans*-2-ene-1,7-dioate aldolase, and succinic semialdehyde dehydrogenase. Enzymes involved in the degradation of tyrosine to fumarate and acetoacetate (e.g., in *Pseudomonas* species) include 4-hydroxyphenylpyruvate dioxygenase, homogentisate 1,2-dioxygenase, maleylacetoacetate isomerase, fumarylacetoacetase and 4-hydroxyphenylacetate. Additional enzymes associated with tyrosine metabolism in different

organisms include 4-chlorophenylacetate-3,4-dioxygenase, aromatic aminotransferase, 5-oxopent-3-ene-1,2,5-tricarboxylate decarboxylase, 2-oxo-hept-3-ene-1,7-dioate hydratase, and 5-carboxymethyl-2-hydroxymuconate isomerase (Ellis, L.B.M. et al. (1999) Nucleic Acids Res. 27:373-376; Wackett, L.P. and Ellis, L.B.M. (1996) J. Microbiol. Meth. 25:91-93; and Schmidt, M. (1996) Amer. Soc. Microbiol. News 62:102).

In humans, acquired or inherited genetic defects in enzymes of the tyrosine degradation pathway may result in hereditary tyrosinemia. One form of this disease, hereditary tyrosinemia 1 (HT1) is caused by a deficiency in the enzyme fumarylacetoacetate hydrolase, the last enzyme in the pathway in organisms that metabolize tyrosine to fumarate and acetoacetate. HT1 is characterized by progressive liver damage beginning at infancy, and increased risk for liver cancer (Endo, F. et al. (1997) J. Biol. Chem. 272:24426-24432).

Expression profiling

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

The discovery of new enzymes, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of autoimmune/inflammatory disorders, infectious disorders, immune deficiencies, disorders of metabolism, reproductive disorders, neurological disorders, cardiovascular disorders, eye disorders, and cell proliferative disorders, including cancer, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of enzymes.

## SUMMARY OF THE INVENTION

The invention features purified polypeptides, enzymes, referred to collectively as "ENZM" and individually as "ENZM-1," "ENZM-2," "ENZM-3," "ENZM-4," "ENZM-5," "ENZM-6," "ENZM-7," "ENZM-8," "ENZM-9," "ENZM-10," "ENZM-11," and "ENZM-12." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ

ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-12.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-12. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:13-24.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid

sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain

reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional ENZM, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional ENZM, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) exposing a sample comprising the

polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional ENZM, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

## BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

## DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

### DEFINITIONS

"ENZM" refers to the amino acid sequences of substantially purified ENZM obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of ENZM. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of ENZM either by directly interacting with ENZM or by acting on components of the biological pathway in which ENZM participates.

An "allelic variant" is an alternative form of the gene encoding ENZM. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in

polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding ENZM include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as ENZM or a polypeptide with at least one functional characteristic of ENZM. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding ENZM, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding ENZM. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent ENZM. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of ENZM is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of ENZM. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of ENZM either by directly interacting with ENZM or by acting on components of the biological pathway in which ENZM participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments

thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind ENZM polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-$NH_2$), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA;

RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic ENZM, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding ENZM or fragments of ENZM may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows

amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

| Original Residue | Conservative Substitution |
|---|---|
| Ala | Gly, Ser |
| Arg | His, Lys |
| Asn | Asp, Gln, His |
| Asp | Asn, Glu |
| Cys | Ala, Ser |
| Gln | Asn, Glu, His |
| Glu | Asp, Gln, His |
| Gly | Ala |
| His | Asn, Arg, Gln, Glu |
| Ile | Leu, Val |
| Leu | Ile, Val |
| Lys | Arg, Gln, Glu |
| Met | Leu, Ile |
| Phe | His, Met, Leu, Trp, Tyr |
| Ser | Cys, Thr |
| Thr | Ser, Val |
| Trp | Phe, Tyr |
| Tyr | His, Phe, Trp |
| Val | Ile, Leu, Thr |

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an

exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of ENZM or the polynucleotide encoding ENZM which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:13-24 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:13-24, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:13-24 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:13-24 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:13-24 and the region of SEQ ID NO:13-24 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-12 is encoded by a fragment of SEQ ID NO:13-24. A fragment of SEQ ID NO:1-12 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-12. For example, a fragment of SEQ ID NO:1-12 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-12. The precise length of a fragment of SEQ ID NO:1-12 and the region of SEQ ID NO:1-12 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer

to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example,

as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for

example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ($T_m$) for the specific sequence at a defined ionic strength and pH. The $T_m$ is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating $T_m$ and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%.

Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., $C_0t$ or $R_0t$ analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of ENZM which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of ENZM which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of ENZM. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of ENZM.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an ENZM may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of ENZM.

"Probe" refers to nucleic acid sequences encoding ENZM, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge

MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing ENZM, nucleic acids encoding ENZM, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In one alternative, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to

another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

**THE INVENTION**

The invention is based on the discovery of new human enzymes (ENZM), the polynucleotides encoding ENZM, and the use of these compositions for the diagnosis, treatment, or prevention of autoimmune/inflammatory disorders, infectious disorders, immune deficiencies, disorders of metabolism, reproductive disorders, neurological disorders, cardiovascular disorders, eye disorders, and cell proliferative disorders, including cancer.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are enzymes. For example, SEQ ID NO:1 is 38% identical, from residue L59 to residue I250, to human cytosolic epoxide hydrolase (GenBank ID g181395) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.4e-35, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains an alphabeta hydrolase fold domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and additional BLAST analyses provide further corroborative evidence that SEQ ID NO:1 is an epoxide hydrolase. In an alternative example, SEQ ID NO:2 is 64% identical, from residue E275 to residue I864, to mouse acetyltransferase Tubedown-1 (GenBank ID g8497318) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 4.4e-210, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains TPR domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from additional BLAST analysis against the PRODOM database provides further corroborative evidence that SEQ ID NO:2 is an acetyltransferase. In an alternative example, SEQ ID NO:3 is 97% identical, from residue M1 to residue K376, to human trans-prenyltransferase (GenBank ID g4732024) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.6e-194, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also contains a polyprenyl synthetase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:3 is a trans-prenyltransferase. In an alternative example, SEQ ID NO:6 is 75% identical, from residue M1 to residue K399, to rat argininosuccinate

synthetase (GenBank ID g55767) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 6.3e-158, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also contains an arginosuccinate synthase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:6 is an argininosuccinate synthetase. In an alternative example, SEQ ID NO:8 is 100% identical, from residue M1 to Y583, to human succinate dehydrogenase flavoprotein subunit (GenBank ID g347134) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:8 also contains a FAD binding domain and a fumarate reductase/succinate dehydrogenase flavoprotein C-terminal domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, PROFILESCAN and other BLAST analyses provide further corroborative evidence that SEQ ID NO:8 is a succinate dehydrogenase flavoprotein subunit. In an alternative example, SEQ ID NO:10 is 34% identical, from residue V23 to residue L380, to rat arylacetamide deacetylase (GenBank ID g5923874) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.0e-51, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Data from BLIMPS analysis provide further corroborative evidence that SEQ ID NO:10 is an esterase. In an alternative example, SEQ ID NO:12 is 65% identical, from residue M35 to residue L326, to human neuropathy target esterase (GenBank ID g2982501) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.4e-95, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:12 also contains a cyclic nucleotide-binding domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLAST analysis provide further corroborative evidence that SEQ ID NO:12 is a neuropathy target esterase. SEQ ID NO:4-5, SEQ ID NO:7, SEQ ID NO:9, and SEQ ID NO:11 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-12 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide

consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:13-24 or that distinguish between SEQ ID NO:13-24 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_$N_1$_$N_2$_*YYYYY*_$N_3$_$N_4$ represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

| Prefix | Type of analysis and/or examples of programs |
|---|---|
| GNN, GFG, ENST | Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK). |
| GBI | Hand-edited analysis of genomic sequences. |
| FL | Stitched or stretched genomic sequences (see Example V). |
| INCY | Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript. |

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses ENZM variants. A preferred ENZM variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the ENZM amino acid sequence, and which contains at least one functional or structural characteristic of ENZM.

The invention also encompasses polynucleotides which encode ENZM. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:13-24, which encodes ENZM. The polynucleotide sequences of SEQ ID NO:13-24, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding ENZM. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding ENZM. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:13-24 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting

of SEQ ID NO:13-24. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of ENZM.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding ENZM. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding ENZM, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding ENZM over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding ENZM. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of ENZM.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding ENZM, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring ENZM, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode ENZM and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring ENZM under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding ENZM or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding ENZM and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode ENZM and ENZM derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell

systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding ENZM or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:13-24 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding ENZM may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve

unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode ENZM may be cloned in recombinant DNA molecules that direct expression of ENZM, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express ENZM.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter ENZM-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such

as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of ENZM, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding ENZM may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, ENZM itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of ENZM, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active ENZM, the nucleotide sequences encoding ENZM or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences

encoding ENZM. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding ENZM. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding ENZM and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding ENZM and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding ENZM. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.)

The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding ENZM. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding ENZM can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding ENZM into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of ENZM are needed, e.g. for the production of antibodies, vectors which direct high level expression of ENZM may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of ENZM. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of ENZM. Transcription of sequences encoding ENZM may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding ENZM may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses ENZM in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma

virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of ENZM in cell lines is preferred. For example, sequences encoding ENZM can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding ENZM is inserted within a marker gene sequence, transformed cells containing sequences encoding ENZM can be identified by the absence of marker gene function. Alternatively,

a marker gene can be placed in tandem with a sequence encoding ENZM under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding ENZM and that express ENZM may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of ENZM using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on ENZM is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding ENZM include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding ENZM, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding ENZM may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode ENZM may be designed to contain signal sequences which

direct secretion of ENZM through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding ENZM may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric ENZM protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of ENZM activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the ENZM encoding sequence and the heterologous protein sequence, so that ENZM may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled ENZM may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, $^{35}$S-methionine.

ENZM of the present invention or fragments thereof may be used to screen for compounds that specifically bind to ENZM. At least one and up to a plurality of test compounds may be screened

for specific binding to ENZM. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of ENZM, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which ENZM binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express ENZM, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing ENZM or cell membrane fractions which contain ENZM are then contacted with a test compound and binding, stimulation, or inhibition of activity of either ENZM or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with ENZM, either in solution or affixed to a solid support, and detecting the binding of ENZM to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

ENZM of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of ENZM. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for ENZM activity, wherein ENZM is combined with at least one test compound, and the activity of ENZM in the presence of a test compound is compared with the activity of ENZM in the absence of the test compound. A change in the activity of ENZM in the presence of the test compound is indicative of a compound that modulates the activity of ENZM. Alternatively, a test compound is combined with an in vitro or cell-free system comprising ENZM under conditions suitable for ENZM activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of ENZM may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding ENZM or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal

models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding ENZM may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding ENZM can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding ENZM is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress ENZM, e.g., by secreting ENZM in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

**THERAPEUTICS**

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of ENZM and enzymes. In addition, examples of tissues expressing ENZM can be found in Table 6. Therefore, ENZM appears to play a role in autoimmune/inflammatory disorders, infectious disorders, immune deficiencies, disorders of metabolism, reproductive disorders, neurological disorders, cardiovascular disorders, eye disorders, and cell proliferative disorders, including cancer. In the treatment of disorders associated with increased ENZM expression or activity, it is desirable to decrease the expression or activity of ENZM. In the treatment of disorders associated with decreased ENZM expression or activity, it is desirable to increase the expression or

activity of ENZM.

Therefore, in one embodiment, ENZM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ENZM. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, and trauma; an infectious disorder such as a viral infection, e.g., caused by an adenovirus (acute respiratory disease, pneumonia), an arenavirus (lymphocytic choriomeningitis), a bunyavirus (Hantavirus), a coronavirus (pneumonia, chronic bronchitis), a hepadnavirus (hepatitis), a herpesvirus (herpes simplex virus, varicella-zoster virus, Epstein-Barr virus, cytomegalovirus), a flavivirus (yellow fever), an orthomyxovirus (influenza), a papillomavirus (cancer), a paramyxovirus (measles, mumps), a picornovirus (rhinovirus, poliovirus, coxsackie-virus), a polyomavirus (BK virus, JC virus), a poxvirus (smallpox), a reovirus (Colorado tick fever), a retrovirus (human immunodeficiency virus, human T lymphotropic virus), a rhabdovirus (rabies), a rotavirus (gastroenteritis), and a togavirus (encephalitis, rubella), and a bacterial infection, a fungal infection, a parasitic infection, a protozoal infection, and a helminthic infection; an immune deficiency, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, and immunodeficiency associated with Cushing's disease; a disorder of metabolism such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism,

hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, a lipid myopathy, a lipodystrophy, a lysosomal storage disease, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria, pseudovitamin D-deficiency rickets; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis, cancer of the breast, fibrocystic breast disease, and galactorrhea, disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; seasonal affective disorder (SAD); akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus,

carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an eye disorder such as ocular hypertension and glaucoma; a disorder of cell proliferation such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia; and a cancer, including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing ENZM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ENZM including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified ENZM in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ENZM including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of ENZM may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ENZM including, but not limited to, those listed above.

In a further embodiment, an antagonist of ENZM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of ENZM. Examples of such disorders include, but are not limited to, those autoimmune/inflammatory disorders, infectious disorders, immune deficiencies, disorders of metabolism, reproductive disorders, neurological

disorders, cardiovascular disorders, eye disorders, and cell proliferative disorders, including cancer described above. In one aspect, an antibody which specifically binds ENZM may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express ENZM.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding ENZM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of ENZM including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of ENZM may be produced using methods which are generally known in the art. In particular, purified ENZM may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind ENZM. Antibodies to ENZM may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with ENZM or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to ENZM have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or

fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of ENZM amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to ENZM may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce ENZM-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for ENZM may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between ENZM and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering ENZM epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for ENZM. Affinity is expressed as an association constant, $K_a$, which is defined as the molar concentration of ENZM-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The $K_a$ determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple ENZM epitopes, represents the average affinity, or avidity, of the antibodies for ENZM. The $K_a$ determined for a preparation of monoclonal antibodies, which are monospecific for a particular ENZM epitope, represents a true measure of affinity. High-affinity antibody preparations with $K_a$ ranging from about $10^9$ to $10^{12}$ L/mole are preferred for use in immunoassays in which the ENZM-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with $K_a$ ranging from about $10^6$ to $10^7$ L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of ENZM, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of ENZM-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding ENZM, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding ENZM. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding ENZM. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995)

9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding ENZM may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in ENZM expression or regulation causes disease, the expression of ENZM from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in ENZM are treated by constructing mammalian expression vectors encoding ENZM and introducing these vectors by mechanical means into ENZM-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of ENZM include, but are not

limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). ENZM may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding ENZM from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to ENZM expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding ENZM under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-

cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding ENZM to cells which have one or more genetic abnormalities with respect to the expression of ENZM. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding ENZM to target cells which have one or more genetic abnormalities with respect to the expression of ENZM. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing ENZM to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to

deliver polynucleotides encoding ENZM to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for ENZM into the alphavirus genome in place of the capsid-coding region results in the production of a large number of ENZM-coding RNAs and the synthesis of high levels of ENZM in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of ENZM into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding ENZM.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides,

corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding ENZM. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding ENZM. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased ENZM expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding ENZM may be therapeutically useful, and in the treatment of disorders associated with decreased ENZM expression or activity, a compound which specifically promotes expression of the polynucleotide encoding ENZM may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in

altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding ENZM is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding ENZM are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding ENZM. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of

Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of ENZM, antibodies to ENZM, and mimetics, agonists, antagonists, or inhibitors of ENZM.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising ENZM or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, ENZM or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example ENZM or fragments thereof, antibodies of ENZM, and agonists, antagonists or inhibitors of ENZM, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the $ED_{50}$ (the dose therapeutically effective in 50% of the population) or $LD_{50}$ (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the

therapeutic index, which can be expressed as the $LD_{50}/ED_{50}$ ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the $ED_{50}$ with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 $\mu$g to 100,000 $\mu$g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind ENZM may be used for the diagnosis of disorders characterized by expression of ENZM, or in assays to monitor patients being treated with ENZM or agonists, antagonists, or inhibitors of ENZM. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for ENZM include methods which utilize the antibody and a label to detect ENZM in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring ENZM, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of ENZM expression. Normal or standard values for ENZM expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to ENZM under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of ENZM expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values.

Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding ENZM may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of ENZM may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of ENZM, and to monitor regulation of ENZM levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding ENZM or closely related molecules may be used to identify nucleic acid sequences which encode ENZM. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding ENZM, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the ENZM encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:13-24 or from genomic sequences including promoters, enhancers, and introns of the ENZM gene.

Means for producing specific hybridization probes for DNAs encoding ENZM include the cloning of polynucleotide sequences encoding ENZM or ENZM derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as $^{32}P$ or $^{35}S$, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding ENZM may be used for the diagnosis of disorders associated with expression of ENZM. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis,

myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, and trauma; an infectious disorder such as a viral infection, e.g., caused by an adenovirus (acute respiratory disease, pneumonia), an arenavirus (lymphocytic choriomeningitis), a bunyavirus (Hantavirus), a coronavirus (pneumonia, chronic bronchitis), a hepadnavirus (hepatitis), a herpesvirus (herpes simplex virus, varicella-zoster virus, Epstein-Barr virus, cytomegalovirus), a flavivirus (yellow fever), an orthomyxovirus (influenza), a papillomavirus (cancer), a paramyxovirus (measles, mumps), a picornovirus (rhinovirus, poliovirus, coxsackie-virus), a polyomavirus (BK virus, JC virus), a poxvirus (smallpox), a reovirus (Colorado tick fever), a retrovirus (human immunodeficiency virus, human T lymphotropic virus), a rhabdovirus (rabies), a rotavirus (gastroenteritis), and a togavirus (encephalitis, rubella), and a bacterial infection, a fungal infection, a parasitic infection, a protozoal infection, and a helminthic infection; an immune deficiency, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, and immunodeficiency associated with Cushing's disease; a disorder of metabolism such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, a lipid myopathy, a lipodystrophy, a lysosomal storage disease, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria, pseudovitamin D-deficiency rickets; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis, cancer of the breast, fibrocystic breast disease, and galactorrhea, disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease,

Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; seasonal affective disorder (SAD); akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced

lung disease, radiation-induced lung disease, and complications of lung transplantation; an eye disorder such as ocular hypertension and glaucoma; a disorder of cell proliferation such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia; and a cancer, including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding ENZM may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered ENZM expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding ENZM may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding ENZM may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding ENZM in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of ENZM, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding ENZM, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from

successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding ENZM may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding ENZM, or a fragment of a polynucleotide complementary to the polynucleotide encoding ENZM, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding ENZM may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding ENZM are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also

useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations. (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641.)

Methods which may also be used to quantify the expression of ENZM include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, ENZM, fragments of ENZM, or antibodies specific for ENZM may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to

generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the

treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for ENZM to quantify the levels of ENZM expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and

should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding ENZM may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific

region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding ENZM on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, ENZM, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between ENZM and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with ENZM, or fragments thereof,

and washed. Bound ENZM is then detected by methods well known in the art. Purified ENZM can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding ENZM specifically compete with a test compound for binding ENZM. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with ENZM.

In additional embodiments, the nucleotide sequences which encode ENZM may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No.60/322,181, U.S. Ser. No.60/315,874, U.S. Ser. No.60/311,447, U.S. Ser. No.60/308,182, U.S. Ser. No.60/293,572, U.S. Ser. No.60/291,544, and U.S. Ser. No.60/283,793 are expressly incorporated by reference herein.

## EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA

purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

**II. Isolation of cDNA Clones**

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

**III. Sequencing and Analysis**

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal

cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS,

DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:13-24. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

**IV. Identification and Editing of Coding Sequences from Genomic DNA**

Putative enzymes were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode enzymes, the encoded polypeptides were analyzed by querying against PFAM models for enzymes. Potential enzymes were also identified by homology to Incyte cDNA sequences that had been annotated as enzymes. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted

sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

**V. Assembly of Genomic Sequence Data with cDNA Sequence Data**

**"Stitched" Sequences**

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

**"Stretched" Sequences**

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions

may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

**VI. Chromosomal Mapping of ENZM Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:13-24 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:13-24 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

**VII. Analysis of Polynucleotide Expression**

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{\text{5 x minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding ENZM are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding ENZM. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

**VIII. Extension of ENZM Encoding Polynucleotides**

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was

synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing $Mg^{2+}$, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

**IX. Identification of Single Nucleotide Polymorphisms in ENZM Encoding Polynucleotides**

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:13-24 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele

frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

**X. Labeling and Use of Individual Hybridization Probes**

Hybridization probes derived from SEQ ID NO:13-24 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 $\mu$Ci of [$\gamma$-$^{32}$P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing $10^7$ counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

**XI. Microarrays**

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The

array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

**Tissue or Cell Sample Preparation**

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/$\mu$l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/$\mu$l RNase inhibitor, 500 $\mu$M dATP, 500 $\mu$M dGTP, 500 $\mu$M dTTP, 40 $\mu$M dCTP, 40 $\mu$M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 $\mu$l 5X SSC/0.2% SDS.

**Microarray Preparation**

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 $\mu$g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water,

and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

**Hybridization**

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 $cm^2$ coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

**Detection**

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source,

although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

**XII. Complementary Polynucleotides**

Sequences complementary to the ENZM-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring ENZM. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of ENZM. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the ENZM-encoding transcript.

**XIII. Expression of ENZM**

Expression and purification of ENZM is achieved using bacterial or virus-based expression systems. For expression of ENZM in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid

promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express ENZM upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of ENZM in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding ENZM by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, ENZM is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from ENZM at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified ENZM obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, and XIX where applicable.

**XIV. Functional Assays**

ENZM function is assessed by expressing the sequences encoding ENZM at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 $\mu$g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 $\mu$g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the

recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of ENZM on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding ENZM and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding ENZM and other genes of interest can be analyzed by northern analysis or microarray techniques.

**XV. Production of ENZM Specific Antibodies**

ENZM substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the ENZM amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-ENZM activity by, for example, binding the peptide or ENZM to a substrate,

blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

**XVI. Purification of Naturally Occurring ENZM Using Specific Antibodies**

Naturally occurring or recombinant ENZM is substantially purified by immunoaffinity chromatography using antibodies specific for ENZM. An immunoaffinity column is constructed by covalently coupling anti-ENZM antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing ENZM are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of ENZM (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/ENZM binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and ENZM is collected.

**XVII. Identification of Molecules Which Interact with ENZM**

ENZM, or biologically active fragments thereof, are labeled with $^{125}$I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled ENZM, washed, and any wells with labeled ENZM complex are assayed. Data obtained using different concentrations of ENZM are used to calculate values for the number, affinity, and association of ENZM with the candidate molecules.

Alternatively, molecules interacting with ENZM are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

ENZM may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

**XVIII. Demonstration of ENZM Activity**

ENZM activity is demonstrated through a variety of specific enzyme assays; some of which are outlined below.

ENZM oxidoreductase activity is measured by the increase in extinction coefficient of NAD(P)H coenzyme at 340 nm for the measurement of oxidation activity, or the decrease in extinction coefficient of NAD(P)H coenzyme at 340 nm for the measurement of reduction activity (Dalziel, K. (1963) J. Biol. Chem. 238:2850-2858). One of three substrates may be used: Asn-βGal, biocytidine, or ubiquinone-10. The respective subunits of the enzyme reaction, for example,

cytochrome $c_1$-b oxidoreductase and cytochrome c, are reconstituted. The reaction mixture contains a)1-2 mg/ml ENZM; and b) 15 mM substrate, 2.4 mM NAD(P)$^+$ in 0.1 M phosphate buffer, pH 7.1 (oxidation reaction), or 2.0 mM NAD(P)H, in 0.1 M $Na_2HPO_4$ buffer, pH 7.4 ( reduction reaction); in a total volume of 0.1 ml. Changes in absorbance at 340 nm ($A_{340}$) are measured at 23.5 °C using a recording spectrophotometer (Shimadzu Scientific Instruments, Inc., Pleasanton, CA). The amount of NAD(P)H is stoichiometrically equivalent to the amount of substrate initially present, and the change in $A_{340}$ is a direct measure of the amount of NAD(P)H produced; $\Delta A_{340}$ = 6620[NADH]. ENZM activity is proportional to the amount of NAD(P)H present in the assay.

Aldo/keto reductase activity of ENZM is proportional to the decrease in absorbance at 340 nm as NADPH is consumed (or increased absorbance if NADPH is produced, i.e., if the reverse reaction is monitored). A standard reaction mixture is 135 mM sodium phosphate buffer (pH 6.2-7.2 depending on enzyme), 0.2 mM NADPH, 0.3 M lithium sulfate, 0.5-2.5 mg ENZM and an appropriate level of substrate. The reaction is incubated at 30 °C and the reaction is monitored continuously with a spectrophotometer. ENZM activity is calculated as mol NADPH consumed / mg of ENZM.

Acyl-CoA dehydrogenase activity of ENZM is measured using an anaerobic electron transferring flavoprotein (ETF) assay. The reaction mixture comprises 50 mM Tris-HCl (pH 8.0), 0.5% glucose, and 50 $\mu$M acyl-CoA substrate (i.e., isovaleryl-CoA) that is pre-warmed to 32 °C. The mixture is depleted of oxygen by repeated exposure to vacuum followed by layering with argon. Trace amounts of oxygen are removed by the addition of glucose oxidase and catalase followed by the addition of ETF to a final concentration of 1 $\mu$M. The reaction is initiated by addition of purified ENZM or a sample containing ENZM and exciting the reaction at 342 nm. Quenching of fluorescence caused by the transfer of electrons from the substrate to ETF is monitored at 496 nm. 1 unit of acyl-CoA dehydrogenase activity is defined as the amount of ENZM required to reduce 1 $\mu$mol of ETF per minute (Reinard, T. et al. (2000) J. Biol. Chem. 275:33738-33743).

Alcohol dehydrogenase activity of ENZM is measured by following the conversion of NAD$^+$ to NADH at 340 nm ($\epsilon_{340}$ = 6.22 $mM^{-1}$ $cm^{-1}$) at 25°C in 0.1 M potassium phosphate (pH 7.5), 0.1 M glycine (pH 10.0), and 2.4 mM NAD$^+$. Substrate (e.g., ethanol) and ENZM are then added to the reaction. The production of NADH results in an increase in absorbance at 340 nm and correlates with the oxidation of the alcohol substrate and the amount of alcohol dehydrogenase activity in the ENZM sample (Svensson, S. (1999) J. Biol. Chem. 274:29712-29719).

Aldehyde dehydrogenase activity of ENZM is measured by determining the total hydrolase + dehydrogenase activity of ENZM and subtracting the hydrolase activity. Hydrolase activity is first determined in a reaction mixture containing 0.05 M Tris-HCl (pH 7.8), 100 mM 2-mercaptoethanol, and 0.5-18 $\mu$M substrate, e.g., 10-HCO-HPteGlu (10-formyltetrahydrofolate; HPteGlu,

tetrahydrofolate) or 10-FDDF (10-formyl-5,8-dideazafolate). Approximately 1$\mu$g of ENZM is added in a final volume of 1.0 ml. The reaction is monitored and read against a blank cuvette, containing all components except enzyme. The appearance of product is measured at either 295 nm for 5,8-dideazafolate or 300 nm for HPteGlu using molar extinction coefficients of $1.89x10^4$ and $2.17x10^4$ for 5,8-dideazafolate and HPteGlu, respectively. The addition of $NADP^+$ to the reaction mixture allows the measurement of both dehydrogenase and hydrolase activity (assays are performed as before). Based on the production of product in the presence of $NADP^+$ and the production of product in the absence of the cofactor, aldehyde dehydrogenase activity is calculated for ENZM. In the alternative, aldehyde dehydrogenase activity is assayed using propanal as substrate. The reaction mixture contains 60 mM sodium pyrophosphate buffer (pH 8.5), 5 mM propanal, 1 mM $NADP^+$, and ENZM in a total volume of 1 ml. Activity is determined by the increase in absorbance at 340 nm, resulting from the generation of NADPH, and is proportional to the aldehyde dehydrogenase activity in the sample (Krupenko, S.A. et al. (1995) J. Biol. Chem. 270:519-522).

6-phosphogluconate dehydrogenase activity of ENZM is measured by incubating purified ENZM, or a composition comprising ENZM, in 120 mM triethanolamine (pH 7.5), 0.1 mM EDTA, 0.5 mM $NADP^+$, and 10-150 $\mu$M 6-phosphogluconate as substrate at 20-25 °C. The production of NADPH is measured fluorimetrically (340 nm excitation, 450 nm emission) and is indicative of 6-phosphogluconate dehydrogenase activity. Alternatively, the production of NADPH is measured photometrically, based on absorbance at 340 nm. The molar amount of NADPH produced in the reaction is proportional to the 6-phosphogluconate dehydrogenase activity in the sample (Tetaud, E. et al. (1999) Biochem. J. 338:55-60).

Ribonucleotide diphosphate reductase activity of ENZM is determined by incubating purified ENZM, or a composition comprising ENZM, along with dithiothreitol, $Mg^{++}$, and ADP, GDP, CDP, or UDP substrate. The product of the reaction, the corresponding deoxyribonucleotide, is separated from the substrate by thin-layer chromatography. The reaction products can be distinguished from the reactants based on rates of migration. The use of radiolabeled substrates is an alternative for increasing the sensitivity of the assay. The amount of deoxyribonucleotides produced in the reaction is proportional to the amount of ribonucleotide diphosphate reductase activity in the sample (note that this is true only for pre-steady state kinetic analysis of ribonucleotide diphosphate reductase activity, as the enzyme is subject to negative feedback inhibition by products) (Nutter, L.M. and Y.-C. Cheng (1984) Pharmac. Ther. 26:191-207).

Dihydrodiol dehydrogenase activity of ENZM is measured by incubating purified ENZM, or a composition comprising ENZM, in a reaction mixture comprising 50 mM glycine (pH 9.0), 2.3 mM $NADP^+$, 8% DMSO, and a trans-dihydrodiol substrate, selected from the group including but not limited to, (±)-trans-naphthalene-1,2-dihydrodiol, (±)-trans-phenanthrene-1,2-dihydrodiol, and (±)-

trans-chrysene-1,2-dihydrodiol. The oxidation reaction is monitored at 340 nm to detect the formation of NADPH, which is indicative of the oxidation of the substrate. The reaction mixture can also be analyzed before and after the addition of ENZM by circular dichroism to determine the stereochemistry of the reaction components and determine which enantiomers of a racemic substrate composition are oxidized by the ENZM (Penning, T.M. (1993) Chemico-Biological Interactions 89:1-34).

Glutathione S-transferase (GST) activity of ENZM is determined by measuring the ENZM catalyzed conjugation of GSH with 1-chloro-2,4-dinitrobenzene (CDNB), a common substrate for most GSTs. ENZM is incubated with 1 mM CDNB and 2.5 mM GSH together in 0.1M potassium phosphate buffer, pH 6.5, at 25 °C. The conjugation reaction is measured by the change in absorbance at 340 nm using an ultraviolet spectrophometer. ENZM activity is proportional to the change in absorbance at 340 nm.

15-oxoprostaglandin 13-reductase (PGR) activity of ENZM is measured following the separation of contaminating 15-hydroxyprostaglandin dehydrogenase (15-PGDH) activity by DEAE chromatography. Following isolation of PGR containing fractions (or using the purified ENZM), activity is assayed in a reaction comprising 0.1 M sodium phosphate (pH 7.4), 1 mM 2-mercaptoethanol, 20 μg substrate (e.g., 15-oxo derivatives of prostaglandins $PGE_1$, $PGE_2$, and $PGE_{2\alpha}$), and 1 mM NADH (or a higher concentration of NADPH). ENZM is added to the reaction which is then incubated for 10 min at 37 °C before termination by the addition of 0.25 ml 2 N NaOH. The amount of 15-oxo compound remaining in the sample is determined by measuring the maximum absorption at 500 nm of the terminated reaction and comparing this value to that of a terminated control reaction that received no ENZM. 1 unit of enzyme is defined as the amount required to catalyze the oxidation of 1 μmol substrate per minute and is proportional to the amount of PGR activity in the sample.

Choline dehydrogenase activity of ENZM is identified by the ability of E. coli, transformed with an ENZM expression vector, to grow on media containing choline as the sole carbon and nitrogen source. The ability of the transformed bacteria to thrive is indicative of choline dehydrogenase activity (Magne Østerås, M. (1998) Proc. Natl. Acad. Sci. USA 95:11394-11399).

ENZM thioredoxin activity is assayed as described (Luthman, M. (1982) Biochemistry 21:6628-6633). Thioredoxins catalyze the formation of disulfide bonds and regulate the redox environment in cells to enable the necessary thiol:disulfide exchanges. One way to measure the thiol:disulfide exchange is by measuring the reduction of insulin in a mixture containing 0.1 M potassium phosphate, pH 7.0, 2 mM EDTA, 0.16 μM insulin, 0.33 mM DTT, and 0.48 mM NADPH. Different concentrations of ENZM are added to the mixture, and the reaction rate is followed by monitoring the oxidation of NADPH at 340 nM.

ENZM transferase activity is measured through assays such as a methyl transferase assay in which the transfer of radiolabeled methyl groups between a donor substrate and an acceptor substrate is measured (Bokar, J.A. et al. (1994) J. Biol. Chem. 269:17697-17704). Reaction mixtures (50 μl final volume) contain 15 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM dithiothreitol, 3% polyvinylalcohol, 1.5 μCi [*methyl*-$^3$H]AdoMet (0.375 μM AdoMet) (DuPont-NEN), 0.6 μg ENZM, and acceptor substrate (0.4 μg [$^{35}$S]RNA or 6-mercaptopurine (6-MP) to 1 mM final concentration). Reaction mixtures are incubated at 30 °C for 30 minutes, then at 65 °C for 5 minutes. The products are separated by chromatography or electrophoresis and the level of methyl transferase activity is determined by quantification of *methyl*-$^3$H recovery.

Aminotransferase activity of ENZM is assayed by incubating samples containing ENZM for 1 hour at 37°C in the presence of 1 mM L-kynurenine and 1 mM 2-oxoglutarate in a final volume of 200 μl of 150 mM Tris acetate buffer (pH 8.0) containing 70 μM PLP. The formation of kynurenic acid is quantified by HPLC with spectrophotometric detection at 330 nm using the appropriate standards and controls well known to those skilled in the art. In the alternative, L-3-hydroxykynurenine is used as substrate and the production of xanthurenic acid is determined by HPLC analysis of the products with UV detection at 340 nm. The production of kynurenic acid and xanthurenic acid, respectively, is indicative of aminotransferase activity (Buchli, R. et al. (1995) J. Biol. Chem. 270:29330-29335).

In another alternative, aminotransferase activity of ENZM is measured by determining the activity of purified ENZM or crude samples containing ENZM toward various amino and oxo acid substrates under single turnover conditions by monitoring the changes in the UV/VIS absorption spectrum of the enzyme-bound cofactor, pyridoxal 5'-phosphate (PLP). The reactions are performed at 25°C in 50 mM 4-methylmorpholine (pH 7.5) containing 9 μM purified ENZM or ENZM containing samples and substrate to be tested (amino and oxo acid substrates). The half-reaction from amino acid to oxo acid is followed by measuring the decrease in absorbance at 360 nm and the increase in absorbance at 330 nm due to the conversion of enzyme-bound PLP to pyridoxamine 5' phosphate (PMP). The specificity and relative activity of ENZM is determined by the activity of the enzyme preparation against specific substrates (Vacca, R. A. et al. (1997) J. Biol. Chem. 272:21932-21937).

ENZM chitinase activity is determined with the fluorogenic substrates 4-methylumbelliferyl chitotriose, methylumbelliferyl chitobiose, or methylumbelliferyl N-acetylglucosamine. Purified ENZM is incubated with 0.5uM substrate at pH 4.0 (0.1M citrate buffer), pH 5.0 (0.1M phosphate buffer), or pH 6.0 (0.1M Tris-HCL). After various times of incubation, the reaction is stopped by the addition of 0.1M glycine buffer, pH 10.4, and the concentration of free methylumbelliferone is determined fluorometrically. Chitinase B from Serratia marcescens may be used as a positive control

(Hakala, B. E. (1993) J. Biol. Chem. 268 (34):25803-25810).

ENZM isomerase activity is determined by measuring 2-hydroxyhepta-2,4-diene,1,7 dioate isomerase (HHDD isomerase) activity, as described by Garrido-Peritierra, A. and Cooper, R.A. (Eur. J. Biochem. (1981)17:581-584). The sample is combined with 5-carboxymethyl-2-oxo-hex-3-ene-1,5, dioate (CMHD), which is the substrate for HHDD isomerase. CMHD concentration is monitored by measuring its absorbance at 246 nm. Decrease in absorbance at 246 nm is proportional to HHDD isomerase activity of ENZM.

ENZM isomerase activity such as peptidyl prolyl *cis/trans* isomerase activity can be assayed by an enzyme assay described by Rahfeld, J. U., et al. (1994) (FEBS Lett. 352: 180-184). The assay is performed at 10 °C in 35 mM HEPES buffer, pH 7.8, containing chymotrypsin (0.5 mg/ml) and ENZM at a variety of concentrations. Under these assay conditions, the substrate, Suc-Ala-Xaa-Pro-Phe-4-NA, is in equilibrium with respect to the prolyl bond, with 80-95% in *trans* and 5-20% in *cis* conformation. An aliquot (2 $\mu$l) of the substrate dissolved in dimethyl sulfoxide (10 mg/ml) is added to the reaction mixture described above. Only the *cis* isomer is a substrate for cleavage by chymotrypsin. Thus, as the substrate is isomerized by ENZM, the product is cleaved by chymotrypsin to produce 4-nitroanilide, which is detected by its absorbance at 390 nm. 4-Nitroanilide appears in a time-dependent and a ENZM concentration-dependent manner.

Alternatively, peptidyl prolyl *cis-trans* isomerase activity of ENZM can be assayed using a chromogenic peptide in a coupled assay with chymotrypsin (Fischer, G. et al. (1984) Biomed. Biochim. Acta 43:1101-1111).

UDP glucuronyltransferase activity of ENZM is measured using a colorimetric determination of free amine groups (Gibson, G.G. and P. Skett (1994) Introduction to Drug Metabolism, Blackie Academic and Professional, London). An amine-containing substrate, such as 2-aminophenol, is incubated at 37 °C with an aliquot of the enzyme in a reaction buffer containing the necessary cofactors (40 mM Tris pH 8.0, 7.5 mM $MgCl_2$, 0.025% Triton X-100, 1 mM ascorbic acid, 0.75 mM UDP-glucuronic acid). After sufficient time, the reaction is stopped by addition of ice-cold 20% trichloroacetic acid in 0.1 M phosphate buffer pH 2.7, incubated on ice, and centrifuged to clarify the supernatant. Any unreacted 2-aminophenol is destroyed in this step. Sufficient freshly-prepared sodium nitrite is then added; this step allows formation of the diazonium salt of the glucuronidated product. Excess nitrite is removed by addition of sufficient ammonium sulfamate, and the diazonium salt is reacted with an aromatic amine (for example, N-naphthylethylene diamine) to produce a colored azo compound which can be assayed spectrophotometrically (at 540 nm, for example). A standard curve can be constructed using known concentrations of aniline, which will form a chromophore with similar properties to 2-aminophenol glucuronide.

Adenylosuccinate synthetase activity of ENZM is measured by synthesis of AMP from IMP.

The sample is combined with AMP. IMP concentration is monitored spectrophotometrically at 248 nm at 23°C (Wang, W. et al. (1995) J. Biol. Chem. 270:13160-13163). The increase in IMP concentration is proportional to ENZM activity.

Alternatively, AMP binding activity of ENZM is measured by combining the sample with $^{32}$P-labeled AMP. The reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to remove unbound label. The radioactivity retained in the gel is proportional to ENZM activity.

In another alternative, xenobiotic carboxylic acid:CoA ligase activity of ENZM is measured by combining the sample with $\gamma$-$^{33}$P-ATP and measuring the formation of $\gamma$-$^{33}$P- pyrophosphate with time (Vessey, D.A. et al. (1998) J. Biochem. Mol. Toxicol. 12:151-155).

Protein phosphatase (PP) activity can be measured by the hydrolysis of P-nitrophenyl phosphate (PNPP). ENZM is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% β-mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH (Diamond, R.H. et al. (1994) Mol. Cell. Biol. 14:3752-62).

Alternatively, acid phosphatase activity of ENZM is demonstrated by incubating ENZM containing extract with 100 µl of 10 mM PNPP in 0.1 M sodium citrate, pH 4.5, and 50 µl of 40 mM NaCl at 37°C for 20 min. The reaction is stopped by the addition of 0.5 ml of 0.4 M glycine/NaOH, pH 10.4 (Saftig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). The increase in light absorbance at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of ENZM in the assay.

In the alternative, ENZM activity is determined by measuring the amount of phosphate removed from a phosphorylated protein substrate. Reactions are performed with 2 or 4 nM ENZM in a final volume of 30 µl containing 60 mM Tris, pH 7.6, 1 mM EDTA, 1 mM EGTA, 0.1% 2-mercaptoethanol and 10 µM substrate, $^{32}$P-labeled on serine/threonine or tyrosine, as appropriate. Reactions are initiated with substrate and incubated at 30° C for 10-15 min. Reactions are quenched with 450 µl of 4% (w/v) activated charcoal in 0.6 M HCl, 90 mM $Na_4P_2O_7$, and 2 mM $NaH_2PO_4$, then centrifuged at 12,000 x g for 5 min. Acid-soluble $^{32}$Pi is quantified by liquid scintillation counting (Sinclair, C. et al. (1999) J. Biol. Chem. 274:23666-23672).

The adenosine deaminase activity of ENZM is determined by measuring the rate of deamination that occurs when adenosine substrate is incubated with ENZM. Reactions are performed with a predetermined amount of ENZM in a final volume of 3.0 ml containing 53.3 mM potassium phosphate and 0.045 mM adenosine. Assay reagents excluding ENZM are mixed in a quartz cuvette and equilibrated to 25° C. Reactions are initiated by the addition of ENZM and are mixed immediately by inversion. The decrease in light absorbance at 265 nm resulting from the hydrolysis of adenosine to inosine is measured using a spectrophotometer. The decrease in the $A_{265\ nm}$ is

recorded for approximately 5 minutes. The decrease in light absorbance is proportional to the activity of ENZM in the assay.

ENZM hydrolase activity is measured by the hydrolysis of appropriate synthetic peptide substrates conjugated with various chromogenic molecules in which the degree of hydrolysis is quantified by spectrophotometric (or fluorometric) absorption of the released chromophore (Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York, NY, pp.25-55). Peptide substrates are designed according to the category of protease activity as endopeptidase (serine, cysteine, aspartic proteases), aminopeptidase (leucine aminopeptidase), or carboxypeptidase (Carboxypeptidase A and B, procollagen C-proteinase).

An assay for carbonic anhydrase activity of ENZM uses the fluorescent pH indicator 8-hydroxypyrene-1,3,6-trisulfonate (pyranine) in combination with stopped-flow fluorometry to measure carbonic anhydrase activity (Shingles, et al. 1997, Anal. Biochem. 252: 190-197). A pH 6.0 solution is mixed with a pH 8.0 solution and the initial rate of bicarbonate dehydration is measured. Addition of carbonic anhydrase to the pH 6.0 solution enables the measurement of the initial rate of activity at physiological temperatures with resolution times of 2 ms. Shingles et al. used this assay to resolve differences in activity and sensitivity to sulfonamides by comparing mammalian carbonic anhydrase isoforms. The fluorescent technique's sensitivity allows the determination of initial rates with a protein concentration as little as 65 ng/ml.

Decarboxylase activity of ENZM is measured as the release of $CO_2$ from labeled substrate. For example, ornithine decarboxylase activity of ENZM is assayed by measuring the release of $CO_2$ from L-[1-$^{14}$C]-ornithine (Reddy, S.G et al. (1996) J. Biol. Chem. 271:24945-24953). Activity is measured in 200 µl assay buffer (50 mM Tris/HCl, pH 7.5, 0.1 mM EDTA, 2 mM dithiothreitol, 5 mM NaF, 0.1% Brij35, 1 mM PMSF, 60 µM pyridoxal-5-phosphate) containing 0.5 mM L-ornithine plus 0.5 µCi L-[1-$^{14}$C]ornithine. The reactions are stopped after 15-30 minutes by addition of 1 M citric acid, and the $^{14}CO_2$ evolved is trapped on a paper disk filter saturated with 20 µl of 2 N NaOH. The radioactivity on the disks is determined by liquid scintillation spectography. The amount of $^{14}CO_2$ released is proportional to ornithine decarboxylase activity of ENZM.

AdoHCYase activity of ENZM in the hydrolytic direction is performed spectroscopically by measuring the rate of the product (homocysteine) formed by reaction with 5,5'-Dithiobis(2-nitrobenzoic acid) (DTNB). To 800 µl of an enzyme solution containing 4.7 µg of ENZM and 4 units of adenosine deaminase in 50 mM potassium phosphate buffer, pH 7.2, containing 1 mM EDTA (buffer A), is added 200 µl of *S*-Adenosyl-L-homocysteine (500 µM) containing 250 µM DTNB in buffer A. The reaction mixture is incubated at 37 °C for 2 minutes. Hydrolytic activity is monitored at 412 nm continuously using a diode array UV spectrophotometer. Enzyme activity is defined as the amount of enzyme that can hydrolyze 1 µmol of *S*-Adenosyl-L-homocysteine/minute (Yuan, C-S et

al. (1996) J. Biol. Chem. 271:28009-28015).

AdoHCYase activity of ENZM can be measured in the synthetic direction as the production of S-adenosyl homocysteine using 3-deazaadenosine as a substrate (Sganga, M.W. et al. supra). Briefly, ENZM is incubated in a 100 μl volume containing 0.1 mM 3-deazaadenosine, 5 mM homocysteine, 20 mM HEPES (pH 7.2). The assay mixture is incubated at 37 °C for 15 minutes. The reaction is terminated by the addition of 10 μl of 3 M perchloric acid. After incubation on ice for 15 minutes, the mixture is centrifuged for 5 minutes at 18,000 x *g* in a microcentrifuge at 4 °C. The supernatant is removed, neutralized by the addition of 1 M potassium carbonate, and centrifuged again. A 50 μl aliquot of supernatant is then chromatographed on an Altex Ultrasphere ODS column (5 μm particles, 4.6 x 250 mm) by isocratic elution with 0.2 M ammonium dihydrogen phosphate (Aldrich) at a flow rate of 1 ml/min. Protein is determined by the bicinchoninic acid assay (Pierce).

Alternatively, AdoHCYase activity of ENZM can be measured in the synthetic direction by a TLC method (Hershfield, M.S. et al. (1979) J. Biol. Chem. 254:22-25). In a preincubation step, 50 μM [8-$^{14}$C]adenosine is incubated with 5 molar equivalents of $NAD^+$ for 15 minutes at 22 °C. Assay samples containing ENZM in a 50 μl final volume of 50 mM potassium phosphate buffer, pH 7.4, 1 mM DTT, and 5 mM homocysteine, are mixed with the preincubated [8-$^{14}$C]adenosine/$NAD^+$ to initiate the reaction. The reaction is incubated at 37 °C, and 1 μl samples are spotted on TLC plates at 5 minute intervals for 30 minutes. The chromatograms are developed in butanol-1/glacial acetic acid/water (12:3:5, v/v) and dried. Standards are used to identify substrate and products under ultraviolet light. The complete spots containing [$^{14}$C]adenosine and [$^{14}$C]SAH are then detected by exposing x-ray film to the TLC plate. The radiolabeled substrate and product are then cut from the chromatograms and counted by liquid scintillation spectrometry. Specific activity of the enzyme is determined from the linear least squares slopes of the product *vs* time plots and the milligrams of protein in the sample (Bethin, K.E. et al. (1995) J. Biol. Chem. 270:20698-20702).

Asparaginase activity of ENZM can be measured in the hydrolytic direction by determining the amount of radiolabeled L-aspartate released from 0.6 mM *N*$^4$-β'-*N*-acetylglucosaminyl-L-asparagine substrate when it is incubated at 25 °C with ENZM in 50 mM phosphate buffer, pH 7.5 (Kaartinen, V. et al. (1991) J. Biol. Chem. 266:5860-5869).

Measurement of acyl CoA acid hydrolase activity of ENZM in the hydrolytic direction is performed spectroscopically by monitoring the appearance of the product (CoASH) formed by reaction of substrate (acyl-CoA) and ENZM with 5,5'-Dithiobis(2-nitrobenzoic acid) (DTNB). The final reaction volume is 1 ml of 0.05 M potassium phosphate buffer, pH 8, containing 0.1 mM DTNB, 20 μg/ml bovine serum albumin, 10 μM of acyl-CoA of different lengths (C6-CoA, C10-CoA, C14-CoA and C18-CoA, Sigma), and ENZM. The reaction mixture is incubated at 22 °C for 7 minutes. Hydrolytic activity is monitored spectrophotometrically by measuring absorbance at 412

nm (Poupon, V. et al. supra).

RNase activity of ENZM can be measured spectrophotometrically by determining the amount of solubilized RNA that is produced as a result of incubation of RNA substrate with ENZM. 5 μl (20 μg) of a 4 mg/ml solution of yeast tRNA (Sigma) is added to 0.8 ml of 40 mM sodium phosphate, pH 7.5, containing ENZM. The reaction is incubated at 25 °C for 15 minutes. The reaction is stopped by addition of 0.5 ml of an ice-cold fresh solution of 20 mM lanthanum nitrate plus 3% perchloric acid. The stopped reaction is incubated on ice for at least 15 min, and the insoluble tRNA is removed by centrifugation for 5 min at 10,000 g. Solubilized tRNA is determined as UV absorbance (260 nm) of the remaining supernatant, with $A_{260}$ of 1.0 corresponding to 40 μg of solubilized RNA (Rosenberg, H.F. et al. (1996) Nucleic Acids Research 24:3507-3513).

RNase P or tRNA splicing endonuclease activity can be determined as the ability of ENZM to cleave $^{32}P$ internally labeled T. thermophila pre-$tRNA^{Gln}$. RNase P or tRNA splicing endonuclease and substrate are added to reaction vessels and reactions are carried out in MBB buffer (50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$) for 1 hour at 37 °C. Reactions are terminated with the addition of an equal volume of sample loading buffer (SLB: 40 mM EDTA, 8 M urea, 0.2% xylene cyanol, and 0.2% bromophenol blue). The reaction products are separated by electrophoresis on 8 M urea, 6% polyacrylamide gels and analyzed using detection instruments and software capable of quantification of the products. One unit of RNase P or tRNA splicing endonuclease activity is defined as the amount of enzyme required to cleave 10% of 28 fmol of T. thermophila pre-$tRNA^{Gln}$ to mature products in 1 hour at 37 °C (True, H.L. et al. (1996) J. Biol. Chem. 271:16559-16566).

Alternatively, cleavage of $^{32}P$ internally labeled substrate tRNA by RNase P or tRNA splicing endonuclease can be determined in a 20 μl reaction mixture containing 30 mM HEPES-KOH (pH 7.6), 6 mM $MgCl_2$, 30 mM KCl, 2 mM DTT, 25 μg/ml bovine serum albumin, 1 unit/μl rRNasin, and 5,000-50,000 cpm of gel-purified substrate RNA. 3.0 μl of RNase P or tRNA splicing endonuclease is added to the reaction mixture, which is then incubated at 37 °C for 30 minutes. The reaction is stopped by guanidinium/phenol extraction, precipitated with ethanol in the presence of glycogen, and subjected to denaturing polyacrylamide gel electrophoresis (6 or 8% polyacrylamide, 7 M urea) and autoradiography (Rossmanith, W. et al. (1995) J. Biol. Chem. 270:12885-12891). The RNase P or tRNA splicing endonuclease activity is proportional to the amount of cleavage products detected.

ENZM activity can be measured by determining the amount of free adenosine produced by the hydrolysis of AMP, as described by Sala-Newby et al. supra. Briefly, ENZM is incubated with AMP in a suitable buffer for 10 minutes at 37 °C. Free adenosine is separated from AMP and measured by reverse phase HPLC.

Alternatively, ENZM activity is measured by the hydrolysis of ADP-ribosylarginine

(Konczalik, P. and J. Moss (1999) J. Biol. Chem. 274:16736-16740). 50 ng of ENZM is incubated with 100 μM ADP-ribosyl-[$^{14}$C]arginine (78,000 cpm) in 50 mM potassium phosphate, pH 7.5, 5 mM dithiothreitol, 10 mM $MgCl_2$ in a final volume of 100 μl. After 1 h at 37° C, 90 μl of the sample is applied to a column (0.5 x 4 cm) of Affi-Gel 601 (boronate) equilibrated and eluted with five 1-ml portions of 0.1 M glycine, pH 9.0, 0.1 M NaCl, and 10 mM $MgCl_2$. Free $^{14}$C-Arg in the total eluate is measured by liquid scintillation counting.

Epoxide hydrolase activity of ENZM can be determined with a radiometric assay utilizing [H$^3$]-labeled *trans*-stilbene oxide (TSO) as substrate. Briefly, ENZM is preincubated in Tris-HCl pH 7.4 buffer in a total volume of 100 μl for 1 minute at 37°C. 1 μl of [H$^3$]-labeled TSO (0.5 μM in EtOH) is added and the reaction mixture is incubated at 37°C for 10 minutes. The reaction mixture is extracted with 200 μl n-dodecane. 50 μl of the aqueous phase is removed for quantification of diol product in a liquid scintillation counter (LSC). ENZM activity is calculated as nmol diol product/min/mg protein (Gill, S.S. et al. (1983) Analytical Biochemistry 131:273-282).

Lysophosphatidic acid acyltransferase activity of ENZM is measured by incubating samples containing ENZM with 1 mM of the thiol reagent 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB), 50 μm LPA, and 50 μm acyl-CoA in 100 mM Tris-HCl, pH 7.4. The reaction is initiated by addition of acyl-CoA, and allowed to reach equilibrium. Transfer of the acyl group from acyl-CoA to LPA releases free CoA, which reacts with DTNB. The product of the reaction between DTNB and free CoA absorbs at 413 nm. The change in absorbance at 413 nm is measured using a spectrophotometer, and is proportional to the lysophosphatidic acid acyltransferase activity of ENZM in the sample.

N-acyltransferase activity of ENZM is measured using radiolabeled amino acid substrates and measuring radiolabel incorporation into conjugated products. ENZM is incubated in a reaction buffer containing an unlabeled acyl-CoA compound and radiolabeled amino acid, and the radiolabeled acyl-conjugates are separated from the unreacted amino acid by extraction into n-butanol or other appropriate organic solvent. For example, Johnson, M. R. et al. (1990; J. Biol. Chem. 266:10227-10233) measured bile acid-CoA:amino acid N-acyltransferase activity by incubating the enzyme with cholyl-CoA and $^3$H-glycine or $^3$H-taurine, separating the tritiated cholate conjugate by extraction into n-butanol, and measuring the radioactivity in the extracted product by scintillation. Alternatively, N-acyltransferase activity is measured using the spectrophotometric determination of reduced CoA (CoASH) described below.

N-acetyltransferase activity of ENZM is measured using the transfer of radiolabel from [$^{14}$C]acetyl-CoA to a substrate molecule (for example, see Deguchi, T. (1975) J. Neurochem. 24:1083-5). Alternatively, a newer spectrophotometric assay based on DTNB reaction with CoASH may be used. Free thiol-containing CoASH is formed during N-acetyltransferase catalyzed transfer of an acetyl group to a substrate. CoASH is detected using the absorbance of DTNB conjugate at 412

nm (De Angelis, J. et al. (1997) J. Biol. Chem. 273:3045-3050). ENZM activity is proportional to the rate of radioactivity incorporation into substrate, or the rate of absorbance increase in the spectrophotometric assay.

Galactosyltransferase activity of ENZM is determined by measuring the transfer of galactose from UDP-galactose to a GlcNAc-terminated oligosaccharide chain in a radioactive assay. (Kolbinger, F. et al. (1998) J. Biol. Chem. 273:58-65.) The ENZM sample is incubated with 14 μl of assay stock solution (180 mM sodium cacodylate, pH 6.5, 1 mg/ml bovine serum albumin, 0.26 mM UDP-galactose, 2 μl of UDP-[$^3$H]galactose), 1 μl of $MnCl_2$ (500 mM), and 2.5 μl of GlcNAcβO-$(CH_2)_8$-$CO_2Me$ (37 mg/ml in dimethyl sulfoxide) for 60 minutes at 37°C. The reaction is quenched by the addition of 1 ml of water and loaded on a C18 Sep-Pak cartridge (Waters), and the column is washed twice with 5 ml of water to remove unreacted UDP-[$^3$H]galactose. The [$^3$H]galactosylated GlcNAcβO-$(CH_2)_8$-$CO_2Me$ remains bound to the column during the water washes and is eluted with 5 ml of methanol. Radioactivity in the eluted material is measured by liquid scintillation counting and is proportional to galactosyltransferase activity of ENZM in the starting sample.

Phosphoribosyltransferase activity of ENZM is measured as the transfer of a phosphoribosyl group from phosphoribosylpyrophosphate (PRPP) to a purine or pyrimidine base. Assay mixture (20 μl) containing 50 mM Tris acetate, pH 9.0, 20 mM 2-mercaptoethanol, 12.5 mM $MgCl_2$, and 0.1 mM labeled substrate, for example, [$^{14}$C]uracil, is mixed with 20 μl of ENZM diluted in 0.1 M Tris acetate, pH 9.7, and 1 mg/ml bovine serum albumin. Reactions are preheated for 1 min at 37 °C, initiated with 10 μl of 6 mM PRPP, and incubated for 5 min at 37 °C. The reaction is stopped by heating at 100°C for 1 min. The product [$^{14}$C]UMP is separated from [$^{14}$C]uracil on DEAE-cellulose paper (Turner, R.J. et al. (1998) J. Biol. Chem. 273:5932-5938). The amount of [$^{14}$C]UMP produced is proportional to the phosphoribosyltransferase activity of ENZM.

ADP-ribosyltransferase activity of ENZM is measured as the transfer of radiolabel from adenine-NAD to agmatine (Weng, B. et al. (1999) J. Biol. Chem. 274:31797-31803). Purified ENZM is incubated at 30°C for 1 hr in a total volume of 300 μl containing 50 mM potassium phosphate (pH. 7.5), 20 mM agmatine, and 0.1 mM [adenine-U-$^{14}$C]NAD (0.05 mCi). Samples (100 μl) are applied to Dowex columns and [$^{14}$C]ADP-ribosylagmatine eluted with 5 ml of water for liquid scintillation counting. The amount of radioactivity recovered is proportional to ADP-ribosyltransferase activity of ENZM.

An ENZM activity assay measures aminoacylation of tRNA in the presence of a radiolabeled substrate. SYNT is incubated with [$^{14}$C]-labeled amino acid and the appropriate cognate tRNA (for example, [$^{14}$C]alanine and $tRNA^{ala}$) in a buffered solution. $^{14}$C-labeled product is separated from free [$^{14}$C]amino acid by chromatography, and the incorporated $^{14}$C is quantified using a scintillation counter. The amount of $^{14}$C-labeled product detected is proportional to the activity of ENZM in this

assay (Ibba, M. et al. (1997) Science 278:1119-1122).

Alternatively, agininosuccinate synthase activity of ENZM is measured based on the conversion of [$^3$H]aspartate to [$^3$H]argininosuccinate. ENZM is incubated with a mixture of [$^3$H]aspartate, citrulline, Tris-HCl (pH 7.5), ATP, $MgCl_2$, KCl, phosphoenolpyruvate, pyruvate kinase, myokinase, and pyrophosphatase, and allowed to proceed for 60 minutes at 37 °C. Enzyme activity was terminated with addition of acetic acid and heating for 30 minutes at 90 °C. [$^3$H]argininosuccinate is separated from un-catalyzed [$^3$H]aspartate by chromatography and quantified by liquid scintillation spectrometry. The amount of [$^3$H]argininosuccinate detected is proportional to the activity of ENZM in this assay (O'Brien, W. E. (1979) Biochemistry 18:5353-5356).

Alternatively, the esterase activity of ENZM is assayed by the hydrolysis of p-nitrophenylacetate (NPA). ENZM is incubated together with 0.1 μM NPA in 0.1 M potassium phosphate buffer (pH 7.25) containing 150 mM NaCl. The hydrolysis of NPA is measured by the increase of absorbance at 400 nm with a spectrophotometer. The increase in light absorbance is proportional to the activity of ENZM (Probst, M.R. et al. (1994) J. Biol. Chem. 269:21650-21656).

**XIX. Identification of ENZM Agonists and Antagonists**

Agonists or antagonists of ENZM activation or inhibition may be tested using the assays described in section XVIII. Agonists cause an increase in ENZM activity and antagonists cause a decrease in ENZM activity.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

| Incyte Project ID | Polypeptide SEQ ID NO: | Incyte Polypeptide ID | Polynucleotide SEQ ID NO: | Incyte Polynucleotide ID | |
|---|---|---|---|---|---|
| 2425607 | 1 | 2425607CD1 | 13 | 2425607CB1 | |
| 2786919 | 2 | 2786919CD1 | 14 | 2786919CB1 | |
| 1801130 | 3 | 1801130CD1 | 15 | 1801130CB1 | |
| 3535146 | 4 | 3535146CD1 | 16 | 3535146CB1 | |
| 1436543 | 5 | 1436543CD1 | 17 | 1436543CB1 | |
| 7491063 | 6 | 7491063CD1 | 18 | 7491063CB1 | |
| 7625645 | 7 | 7625645CD1 | 19 | 7625645CB1 | |
| 5730123 | 8 | 5730123CD1 | 20 | 5730123CB1 | |
| 7481031 | 9 | 7481031CD1 | 21 | 7481031CB1 | |
| 7491216 | 10 | 7491216CD1 | 22 | 7491216CB1 | |
| 71624817 | 11 | 71624817CD1 | 23 | 71624817CB1 | |
| 6945964 | 12 | 6945964CD1 | 24 | 6945964CB1 | |

# Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|---|---|---|---|---|
| 1 | 2425607CD1 | g181395 | 3.40E-35 | [Homo sapiens] cytosolic epoxide hydrolase (Beetham,J.K. et al. (1993) Arch. Biochem. Biophys. 305 (1), 197-201) |
| 2 | 2786919CD1 | g8497318 | 4.40E-210 | [Mus musculus] acetyltransferase Tubedown-1 (Gendron,R.L. et al. (2000) Dev. Dyn. 218 (2), 300-315) |
| 3 | 1801130CD1 | g4732024 | 5.60E-194 | [Homo sapiens] trans-prenyltransferase |
| 4 | 3535146CD1 | g307178 | 3.90E-296 | [Homo sapiens] MDMCSF (EC 1.5.1.5; EC 3.5.4.9; EC 6.3.4.3) (Hum,D.W. et al. (1988) J. Biol. Chem. 263 (31), 15946-15950) |
| 5 | 1436543CD1 | g9951004 | 8.10E-32 | [Pseudomonas aeruginosa] tRNA pseudouridine 55 synthase (Stover,C.K. et al. (2000) Nature 406 (6799), 959-964) |
| 6 | 7491063CD1 | g55767 | 6.30E-158 | [Rattus norvegicus] argininosuccinate synthetase (AA 1-412) (Surh,L.C. et al. (1988) Nucleic Acids Res. 16 (19), 9352) |
| 7 | 7625645CD1 | g7226108 | 8.00E-80 | [fl][Neisseria meningitidis MC58] pantoate--beta-alanine ligase (Tettelin,H. et al. (2000) Science 287 (5459), 1809-1815) |
| 8 | 5730123CD1 | g347134 | 0.00E+00 | [Homo sapiens] succinate dehydrogenase flavoprotein subunit (Morris,A.A. et al. (1994) Biochim. Biophys. Acta 1185 (1), 125-128) |
| 9 | 7481031CD1 | g2982501 | 0.00E+00 | [Homo sapiens] neuropathy target esterase (Lush,M.J. et al. (1998) Biochem. J. 332 (Pt 1), 1-4) |
| 10 | 7491216CD1 | g5923874 | 5.00E-52 | [fl][Rattus norvegicus] arylacetamide deacetylase (Trickett,J.I. et al. (2001) J. Biol. Chem. 276 (43), 39522-39532) |
| 11 | 71624817CD1 | g13421401 | 1.20E-37 | [Caulobacter crescentus] MutT/nudix family protein (Nierman,W.C. et al. (2001) Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141) |
| 12 | 6945964CD1 | g2982501 | 2.40E-95 | [Homo sapiens] neuropathy target esterase (Lush,M.J. et al. (1998) Biochem. J. 332 (Pt 1), 1-4) |

# Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 1 | 2425607CD1 | 363 | S51 S232 T14 T258 T359 Y161 | | signal_cleavage: M1-C26 | SPSCAN |
| | | | | | Signal Peptide: M1-C26, M12-C28, M12-S33, M1-S33 | HMMER |
| | | | | | alpha/beta hydrolase fold: Y121-L355 | HMMER_PFAM |
| | | | | | Transmembrane domain: M12-L40 C164-I191 N-terminus is non-cytosolic | TMAP |
| | | | | | Alpha/beta hydrolase fold signature PR00111: E120-D135, L166-I179, A180-I193, T300-E314 | BLIMPS_PRINTS |
| | | | | | Epoxide hydrolase signature PR00412: H100-K118, E120-D135, L166-I179, A180-I193, L332-F354 | BLIMPS_PRINTS |
| | | | | | Prolyl aminopeptidase (S33) family signature PR00793: M96-E104, V123-T134, L166-A180 | BLIMPS_PRINTS |
| | | | | | PROTEIN HYDROLASE TRANSFERASE PUTATIVE ESTERASE BIOSYNTHESIS EPOXIDE ACYLTRANSFERASE LIPASE SYNTHASE PD000150: S119-N194 | BLAST_PRODOM |
| | | | | | do HYDROLASE; TROPINESTERASE; HYDROXY; DEHYDROGENASE; DM00312 \|P34913\|235-410: T70-F243 \|P34914\|234-409: T70-F243 \|Q01398\|4-174: G80-F236 \|P07383\|12-189: V74-N194 | BLAST_DOMO |

# Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 2 | 2786919CD1 | 864 | S92 S271 S352 S360 S474 S560 S588 S694 S748 S777 S818 T146 T272 T292 T328 T336 T347 T406 T534 T787 T792 Y26 Y260 | N317 N575 N586 N721 N757 N815 N852 | TPR Domain: H80-N113, I408-D441, L374-L407, G46-S79, L224-N257 | HMMER_PFAM |
| | | | | | NTERMINAL ACETYLTRANSFERASE TRANSFERASE AMINOTERMINAL ALPHA AMINO ACYLTRANSFERASE ACETYLATION PD156409: L250-Q522 | BLAST_PRODOM |
| | | | | | Leucine zipper pattern: L373-L394 | MOTIFS |
| 3 | 1801130CD1 | 376 | S78 S122 S148 S149 S235 S275 S316 T41 T224 T284 | N114 N185 | Polyprenyl synthetase: E81-Q342 | HMMER_PFAM |
| | | | | | Transmembrane domain: N157-N185 N-terminus is cytosolic | TMAP |
| | | | | | Polyprenyl synthetases p BL00723: K93-A103, D141-T155, R216-V241, V260-K282 | BLIMPS_BLOCKS |
| | | | | | Polyprenyl synthetases signatures: A238-G291 | PROFILESCAN |
| | | | | | PYROPHOSPHATE SYNTHASE SYNTHETASE TRANSFERASE BIOSYNTHESIS ISOPRENE GERANYLTRANSTRANSFERASE DIPHOSPHATE GERANYLGERANYL FARNESYL PD000572: N185-R313, L82-E215 | BLAST_PRODOM |

# Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| | | | | | POLYPRENYL SYNTHETASES<br>DM00371\|P51268\|5-239: S86-G291<br>DM00371\|P18900\|92-388: R201-A293, N114-L207<br>DM00371\|P31171\|5-239: L82-G291<br>DM00371\|P31114\|32-263: L82-G291 | BLAST_DOMO |
| | | | | | Polyprenyl synthetases signature 1: L138-G152 | MOTIFS |
| | | | | | Polyprenyl synthetases signature 2: V260-D272 | MOTIFS |
| 4 | 3535146CD1 | 978 | S136 S166 S245 S271 S297 S304 S333 S361 S401 S621 S636 S850 T370 T512 T514 T552 T554 T588 T594 T644 T743 T806 T969 Y414 | N125 N290 N546 N689 | Formate--tetrahydrofolate ligase: V359-F978 | HMMER_PFAM |
| | | | | | Tetrahydrofolate dehydrogenase/cyclohydrolase: A119-A228, V78-D108, K253-V288 | HMMER_PFAM |
| | | | | | Transmembrane domain: Q601-L629, N725-K748, I785-F802, I928-T952 N-terminus is cytosolic | TMAP |
| | | | | | Formate—tetrahydrofolate ligase BL00721: P455-H509, A607-V643, D647-K700, G706-P754, E769-I810, H830-A842, I872-G922, V923-D968, A396-K429 | BLIMPS_BLOCKS |
| | | | | | Tetrahydrofolate dehydrogenase/cyclohydrolase BL00766: S135-V182, S201-M248, I266-F295 | BLIMPS_BLOCKS |

# Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 4 | | | | | Tetrahydrofolate dehydrogenase/cyclohydrolase family signature PR00085: P98-E120, S136-N163, V169-L190, D265-D294, V216-L236 | BLIMPS_PRINTS |
| | | | | | SYNTHASE C1TETRAHYDROFOLATE FORMYLTETRAHYDROFOLATE C1THF METHYLENETETRAHYDROFOLATE DEHYDROGENASE METHENYLTETRAHYDROFOLATE CYCLOHYDROLASE FORMATETETRAHYDROFOLATE LIGASE PD003089: V359-F978, A204-G227 | BLAST_PRODOM |
| | | | | | FORMATETETRAHYDROFOLATE LIGASE FORMYLTETRAHYDROFOLATE SYNTHETASE FHS FTHFS ONECARBON METABOLISM ATPBINDING LIGH PD150859: S361-E510 | BLAST_PRODOM |
| | | | | | METHYLENETETRAHYDROFOLATE DEHYDROGENASE CYCLOHYDROLASE METHENYLTETRAHYDROFOLATE INCLUDES: SYNTHASE BIFUNCTIONAL PROTEIN FOLD ONECARBON PD002300: V75-N290, G392-A442 | BLAST_PRODOM |
| | | | | | FORMATE--TETRAHYDROFOLATE LIGASE DM03172\|P11586\|301-933: R345-F978 DM03172\|A43350\|7-637: K352-L977 DM03172\|P07245\|307-945: L351-F978 DM03172\|P13419\|1-555: L564-F978, S361-T554, L173-V225 | BLAST_DOMO |

# Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| | | | | | Formate--tetrahydrofolate ligase signature 1: G462-Y472 | MOTIFS |
| | | | | | Formate--tetrahydrofolate ligase signature 2: V741-G752 | MOTIFS |
| 5 | 1436543CD1 | 349 | S13 S36 S146 S320 T50 T81 T105 T111 T150 T165 T178 T284 T296 T335 Y174 | | Signal Peptides: M31-A54, M31-T50 | HMMER |
| | | | | | Signal cleavage: M31-A73 | SPSCAN |
| | | | | | TruB family pseudouridylate synthase (N terminal domain): E103-I255 | HMMER_PFAM |
| | | | | | TRNA SYNTHASE PROTEIN PSEUDOURIDINE PSEUDOURIDYLATE LYASE PROCESSING PSI55 URACIL HYDROLYASE PD002765: L68-Y254 | BLAST_PRODOM |
| | | | | | PSEUDOURIDINE; TRNA; SYNTHASE; P35 DM05024\|P45142\|1-286: L68-K319 DM05024\|P48567\|32-362: K90-L235 DM05024\|P09171\|1-291: L68-E286 | BLAST_DOMO |
| 6 | 7491063CD1 | 399 | S56 S228 S383 T92 T167 T245 T397 Y156 Y311 | N145 | Arginosuccinate synthase: V8-L392 | HMMER_PFAM |
| | | | | | Argininosuccinate synthase proteins BL00564: M7-D43, Y80-Q100, K105-G149, P211-S228, L241-I286, K329-Y359 | BLIMPS_BLOCKS |
| | | | | | Argininosuccinate synthase signatures: Y90-F143 | PROFILESCAN |

# Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| | | | | | ARGININOSUCCINATE SYNTHASE CITRULLINE-ASPARTATE LIGASE ARGININE BIOSYNTHESIS ATP-BINDING UREA CYCLE DISEASE PD003544: K114-H390, L9-I266 | BLAST_PRODOM |
| | | | | | ARGININOSUCCINATE SYNTHASE DM01454\|P14568\|1-411: M1-K399 DM01454\|P22768\|1-416: S3-T397 DM01454\|P13256\|1-396: V8-K395 DM01454\|P13257\|1-395: M7-K395 | BLAST_DOMO |
| 7 | 7625645CD1 | 278 | S181 S188 S241 T6 T70 T171 | | Pantoate-beta-alanine ligase: M1-I274 | HMMER_PFAM |
| | | | | | PANTOTHENATE SYNTHETASE LIGASE PANTOATEBETAALANINE PANTOATE ACTIVATING ENZYME PROTEIN PUTATIVE PD006306: A81-R192, M1-P96 | BLAST_PRODOM |
| | | | | | do BETA; PANTOATE; LIGASE; DM03293 \|P31663\|1-282: M1-N271 \|Q09673\|1-282: S18-E273 \|P52998\|1-285: N19-I274 \|P40459\|40-344: E9-A261 | BLAST_DOMO |
| 8 | 5730123CD1 | 583 | S12 S44 S161 S177 S208 S272 S321 S456 S470 S474 S505 S547 T126 T545 T553 T567 Y169 Y525 | | FAD binding domain: R162-S492 | HMMER_PFAM |
| | | | | | Fumarate reductase/succinate dehydrogenase flavoprotein C-terminal domain: N487-Y583 | HMMER_PFAM |

# Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 8 | | | | | Fumarate reductase / succinate dehydrogenase FAD-binding site proteins BL00504: G434-R465, D63-G84, T87-N113, P147-R171, R188-Y205, T273-G323, F390-N416 | BLIMPS_BLOCKS |
| | | | | | Fumarate reductase / succinate dehydrogenase FAD-binding site: A77-K128 | PROFILESCAN |
| | | | | | FAD-dependent pyridine nucleotide reductase signature PR00368: D63-F85, T252-R261, I389-T396 | BLIMPS_PRINTS |
| | | | | | Pyridine nucleotide disulphide reductase class-I signature PR00411: D63-F85, T252-R261, I389-T396 | BLIMPS_PRINTS |
| | | | | | Flavin-containing amine oxidase signature PR00757: D63-E82, T213-V234 | BLIMPS_PRINTS |
| | | | | | FLAVOPROTEIN SUBUNIT OXIDOREDUCTASE FAD DEHYDROGENASE SUCCINATE ELECTRON TRANSPORT PROTEIN FUMARATE PD001219: G157-R476 | BLAST_PRODOM |
| | | | | | FLAVOPROTEIN SUBUNIT OXIDOREDUCTASE DEHYDROGENASE SUCCINATE FAD ELECTRON TRANSPORT FUMARATE REDUCTASE PD149640: V90-S161 | BLAST_PRODOM |
| | | | | | SUCCINATE DEHYDROGENASE UBIQUINONE FLAVOPROTEIN SUBUNIT PRECURSOR FP OF COMPLEX II PD040610: M1-S52 | BLAST_PRODOM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 8 | | | | | FLAVOPROTEIN SUBUNIT SUCCINATE DEHYDROGENASE OXIDOREDUCTASE FAD ELECTRON TRANSPORT TRICARBOXYLIC ACID PD149830: W541-Y583 | BLAST_PRODOM |
| | | | | | DEHYDROGENASE; SUCCINATE; FLAVOPROTEIN; YJL045W; DM08853 \|P47052\|44-349: D59-V364 \|S56817\|44-349: D59-V364 | BLAST_DOMO |
| | | | | | FUMARATE REDUCTASE / SUCCINATE DEHYDROGENASE FAD-BINDING SI DM00639 \|P31039\|296-613: F295-K517 \|Q00711\|254-605: Y263-K517 | BLAST_DOMO |
| | | | | | Fumarate reductase / succinate dehydrogenase FAD-binding site: R97-G106 | MOTIFS |
| 9 | 7481031CD1 | 1032 | S25 S49 S75 S133 S152 S160 S184 S228 S297 S333 S392 S433 S494 S604 S823 S894 S936 S950 S959 S1003 S1010 S1028 T257 T510 T526 T547 T914 T977 Y717 Y832 Y885 | N443 N690 N729 N948 | Signal Peptide: M586-S604 | HMMER |
| | | | | | Cyclic nucleotide-binding domain: A306-Q399, R189-T283 | HMMER_PFAM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| | | | | | PROTEIN INTERGENIC REGION TRANSMEMBRANE CHROMOSOME PUTATIVE TWO DOMAIN MEMBRANE CY9C4.03C NEUROPATHY PD004383: P623-G908 | BLAST_PRODOM |
| | | | | | PROTEIN CHROMOSOME NEUROPATHY TARGET ESTERASE OGG1CNA2 INTERGENIC REGION TRANSMEMBRANE ZK370.4 PD017507: D436-H617 | BLAST_PRODOM |
| | | | | | HYPOTHETICAL 100.1 KD PROTEIN M110.7 IN CHROMOSOME II (related to Drosophila Swiss Cheese Mutant Q9W3M0 DROME) PD039093: E230-V400 | BLAST_PRODOM |
| | | | | | NEUROPATHY TARGET ESTERASE PD122528: V915-S1028 | BLAST_PRODOM |
| | | | | | Cell attachment sequence: R355-D357 | MOTIFS |
| 10 | 7491216CD1 | 407 | S127 S262 S393 T110 Y299 | | signal_cleavage: M1-S16 | SPSCAN |
| | | | | | Signal Peptides: M1-F15, M1-S16, M1-G18, M1-T20 | HMMER |
| | | | | | Extracellular domains: M1-D3, E61-L407 Transmembrane domains: L4-S26, L43-F60 Intracellular domains: H27-K42 | TMHMMER |
| | | | | | Carboxylesterases type-B serine proteins BL00122: P113-G123, F159-A199 | BLIMPS_BLOCKS |
| | | | | | Lipolytic enzymes "G-D-X-G" family, histidine BL01173: I115-S127, V146-H172, R186-A199 | BLIMPS_BLOCKS |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|---|---|---|---|---|---|---|
| 11 | 71624817CD1 | 352 | S72 S134 S183 S237 S301 T176 T241 T279 | N320 | signal_cleavage: M1-A63 | SPSCAN |
| | | | | | MutT-like domain: Q197-G321 | HMMER_PFAM |
| | | | | | mutT domain proteins BL00893: L228-L252 | BLIMPS_BLOCKS |
| | | | | | MutT domain signature PR00502: Y225-E239, E239-V254 | BLIMPS_PRINTS |
| | | | | | PROTEIN F13H10.2 RPB9ALG2 INTERGENIC REGION PD069978: S207-L309 | BLAST_PRODOM |
| | | | | | MUTT DOMAIN DM04276 P32664\|2-256: I136-V305 P44710\|1-261: E126-G313 P53164\|79-360: K178-Y315 | BLAST_DOMO |
| | | | | | Leucine zipper pattern: L131-L152, L138-L159 | MOTIFS |
| | | | | | mutT domain signature: G230-E249 | MOTIFS |
| 12 | 6945964CD1 | 385 | S141 S243 S259 T93 T98 T114 T148 | | Signal Peptide: M35-V50; M35-A67 | HMMER |
| | | | | | Cyclic nucleotide-binding domain: H185-E285 | HMMER_PFAM |
| | | | | | Cytosolic domain: R59-E385 Transmembrane domain: L36-F58 Non-cytosolic domain: M1-M35 | TMHMMER |
| | | | | | NEUROPATHY TARGET ESTERASE PD095112: | BLAST_PRODOM |

# Table 4

| Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length | Sequence Fragments |
|---|---|
| 13/2425607CB1/ 1437 | 1-278, 85-563, 247-563, 264-316, 304-927, 304-935, 768-944, 769-944, 770-944, 796-1161, 797-1042, 817-944, 868-1112, 868-1356, 871-1409, 904-1319, 967-1426, 1041-1428, 1043-1428, 1071-1356, 1074-1428, 1263-1412, 1263-1437, 1268-1347, 1334-1426 |
| 14/2786919CB1/ 3269 | 1-590, 1-648, 1-817, 3-470, 25-491, 299-2302, 338-610, 343-938, 382-948, 404-666, 424-895, 435-695, 537-817, 537-1063, 573-850, 588-927, 595-1132, 796-1411, |
| | 891-1467, 1054-1502, 1098-1351, 1271-1526, 1275-1502, 1294-1502, 1305-1502, 1332-1502, 1349-2075, 1433-1958, 1488-1727, 1488-1786, 1784-1929, 1794-2101, |
| | 1840-2061, 1990-2089, 1997-2101, 2021-2296, 2081-2313, 2090-2728, 2103-2774, 2123-2774, 2142-2395, 2165-2400, 2180-2830, 2234-2491, 2247-2774, 2488-2773, |
| | 2518-2787, 2522-2774, 2540-2774, 2540-2775, 2540-2783, 2540-3049, 2654-2898, 2655-3269, 2681-3253, 2708-2918, 2725-3179, 2770-3042, 2777-3269, 2781-3028, 2781-3054, 2781-3064, 2781-3067, 2781-3203, 2781-3269 |
| 15/1801130CB1/ 1534 | 1-732, 372-625, 378-478, 378-568, 378-583, 378-615, 378-617, 378-657, 378-666, 378-686, 378-718, 378-732, 379-732, 405-732, 411-649, 411-732, 441-618, 447-1092, |
| | 474-695, 474-732, 499-682, 499-732, 505-732, 525-732, 554-732, 594-1115, 689-1234, 696-1023, 712-996, 727-1117, 727-1469, 739-1354, 757-1035, 772-1290, |
| | 822-1376, 823-1082, 858-1118, 858-1365, 858-1420, 880-1187, 886-1354, 912-1151, 913-1170, 913-1172, 936-1534, 957-1470 |
| 16/3535146CB1/ 3400 | 1-783, 220-567, 223-789, 223-843, 239-693, 241-409, 241-588, 242-588, 252-386, 263-722, 282-692, 285-1142, 291-461, 291-466, 292-796, 299-670, 303-529, 303-706, 303-794, 308-553, 314-590, 316-601, 433-992, 463-636, 626-1230, 626-1283, 818-1331, 851-1473, 957-1444, 1044-1312, 1072-1313, 1072-1426, 1122-1529, 1142-1757, 1168-1375, 1169-1304, 1201-1608, 1254-1494, 1322-2090, 1465-1700, 1483-2061, 1483-2064, 1486-2064, 1495-2064, 1524-2109, 1538-2185, 1550-1767, 1556-2047, 1577-2243, 1587-2128, 1591-2362, 1746-2199, 1748-1869, 1756-1995, 1756-2296, 1779-2369, 1780-2044, 1790-2249, 1790-2272, 1790-2299, 1790-2343, 1790-2390, 1790-2403, 1790-2405, 1790-2406, 1793-2408, 1808-2657, 1827-2464, 1832-2083, 1832-2456, 1832-2512, 1834-3400, 1854-2506, 1926-2170, 1930-2572, 1978-2251, 1978-2262, 1978-2485, 2066-2312, 2084-2190, 2090-2748, 2191-2353, 2365-3059, 2369-3079, 2373-3194, 2376-3050, 2383-3018, 2395-2835, 2457-3082, 2527-2933, 2527-3020, 2527-3035, 2527-3083, 2527-3139, 2527-3249, 2527-3316, 2531-3083, 2559-3033, 2575-3276, 2596-2857, 2606-2860, 2624-2912, 2641-2907, 2649-3371, 2674-2897, 2677-3350, 2679-3194, 2688-2943, 2702-2932, |
| | 2712-3374, 2737-3001, 2744-3116, 2749-3386, 2760-3383, 2766-3029, 2766-3048, 2767-3087, 2779-3400, 2842-2991, 2872-2962, 2913-2933 |

# Table 4

| Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length | Sequence Fragments |
|---|---|
| 17/1436543CB1/ 3395 | 1-207, 1-450, 15-219, 35-663, 153-313, 153-543, 153-637, 154-650, 157-429, 162-393, 162-530, 162-849, 173-589, 174-306, 174-312, 174-543, 179-628, 192-496, 204-892, 208-421, 208-435, 208-510, 241-613, 471-1017, 491-865, 525-933, 525-1025, 578-839, 584-652, 622-923, 688-944, 696-1301, 738-1361, 746-1183, 798-1077, 836-1409, 886-1380, 887-1380, 947-1263, 972-1259, 972-1477, 989-1590, 1048-1321, 1050-1395, 1071-1726, 1280-1483, 1317-1770, 1324-1571, 1324-1858, 1373-1609, 1535-1801, 1623-1884, 1623-2197, 1672-1817, 1731-1982, 1777-2328, 2003-2218, 2003-2476, 2010-2480, 2064-2339, 2089-2334, 2106-2586, 2149-2386, 2156-2677, 2227-2614, 2254-2535, 2254-2538, 2255-2526, 2286-2527, 2286-2883, 2288-2440, 2350-2638, 2431-2914, 2433-2725, 2457-2640, 2506-2696, 2516-2780, 2540-2795, 2567-3169, 2569-2827, 2616-2835, 2616-3286, 2637-2884, 2656-2919, 2672-3350, 2676-3352, 2682-3362, 2721-3357, 2728-3004, 2760-2910, 2772-3358, 2809-3085, 2817-3386, 2857-3395, 2862-3390, 2862-3395, 2871-3153, 2875-3361, 2876-3343, 2876-3347, 2876-3355, 2889-3382, 2892-3366, 2914-3365, 2915-3369, 2915-3389, 2918-3366, 2919-3366, 2919-3369, 2922-3375, |
| | 2923-3171, 2923-3345, 2923-3367, 2923-3380, 2925-3366, 2927-3367, 2935-3198, 2937-3366, 2953-3350, 2955-3392, 2962-3361, 2966-3196, 2968-3366, |
| | 2969-3381, 2976-3361, 2977-3366, 2977-3386, 2983-3367, 2984-3366, 3015-3249, 3015-3364, 3031-3369, 3032-3367, 3043-3370, 3063-3384, 3065-3341, 3080-3324, |
| | 3081-3369, 3111-3371, 3114-3366, 3120-3382, 3131-3364, 3138-3361, 3153-3366, 3176-3388, 3244-3366, 3244-3393, 3248-3375 |
| 18/7491063CB1/ 2283 | 1-1132, 477-701, 991-1843, 1062-1657, 1284-1880, 1384-2017, 1385-1606, 1385-1792, 1385-1879, 1385-1884, 1385-1888, 1385-1891, 1385-1907, 1385-1916, 1385-1974, 1385-2036, 1387-2283 |
| 19/7625645CB1/ 1193 | 1-256, 1-279, 1-646, 5-621, 62-776, 135-442, 157-442, 430-1187, 430-1193, 521-1193, 587-1193, 598-1193, 601-1193, 611-1193 |
| 20/5730123CB1/ 2092 | 1-654, 33-129, 33-288, 33-302, 33-318, 33-320, 33-330, 33-364, 33-400, 33-409, 33-451, 33-454, 33-481, 33-551, 33-562, 33-635, 35-259, 35-469, 37-291, 39-279, 39-297, 39-318, 39-345, 40-656, 42-251, 43-186, 43-248, 43-315, 44-296, 44-310, 44-315, 45-164, 45-274, 45-279, 45-297, 45-306, 45-339, 45-453, 45-471, 45-621, 45-722, 46-207, 46-224, 46-228, 46-246, 46-280, 46-308, 46-337, 46-350, 46-489, 49-296, 49-377, 50-201, 50-330, 52-256, 52-329, 57-316, 58-147, 58-257, 58-325, 58-329, 58-332, 59-201, 60-733, 86-698, 262-885, 348-770, 350-617, 373-621, 387-645, 397-834, 397-844, 405-614, 405-705, 406-897, 408-688, 409-711, 427-677, 427-995, 428-726, 430-693, 436-885, 447-1151, 447-1223, 458-909, 474-982, 478-756, 481-976, 492-753, 492-1041, 504-771, 505-1074, 506-828, 521-784, 522-774, 524-1017, 525-970, 530-837, 536-806, 536-856, 537-794, 546-1057, 558-924, 561-858, 570-895, 582-863, 583-795, 583-895, 610-766, 610-818, 610-844, 610-855, 610-894, 610-905, 613-1295, 614-722, 614-911, 618-867, 618-905, 618-1029, 624-963, 627-969, 634-1073, 640-968, |

# Table 4

| Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length | Sequence Fragments |
|---|---|
| | 641-884, 659-899, 669-929, 672-925, 679-905, 688-961, 688-974, 688-988, 688-993, 690-914, 691-978, 704-945, 714-959, 714-1236, 721-939, 722-1095, 724-967, 724-968, |
| | 734-981, 736-1028, 740-995, 740-1018, 749-988, 755-1045, 758-1003, 760-1007, 767-987, 767-1013, 983-1061, 1058-1744, 1058-1748, 1058-1776, 1058-1784, |
| | 1058-1790, 1058-1818, 1058-1840, 1058-1842, 1058-1856, 1063-1731, 1072-1818, 1106-1838, 1226-2083, 1304-2087, 1318-1976, 1431-2087, 1588-2092 |
| 21/7481031CB1/ 4213 | 1-581, 105-130, 105-132, 105-133, 105-163, 105-164, 105-165, 105-194, 105-204, 105-227, 105-234, 105-236, 105-242, 105-265, 105-278, 105-301, 105-349, 105-356, 105-360, 105-361, 106-336, 109-360, 111-359, 113-360, 117-360, 140-360, 163-355, 186-360, |
| | 194-360, 199-360, 202-360, 221-360, 232-360, 261-360, 292-357, 330-360, 491-1286, 637-3078, 830-1350, 835-1447, 837-1404, 838-1512, 842-1451, 881-1163, 899-1026, 982-1286, 994-1419, 1000-1439, 1000-1440, 1000-1450, 1000-1488, 1000-1514, |
| | 1000-1521, 1000-1551, 1000-1552, 1000-1592, 1000-1615, 1000-1622, 1000- |
| | 1689, 1000-1785, 1035-1624, 1041-1630, 1087-1653, 1094-1858, 1114-1419, 1140-1803, 1204-1520, 1258-1973, 1274-1513, 1275-1518, 1277-1518, 1278-1518, 1278-1975, |
| | 1285-1553, 1319-1582, 1319-1841, 1319-1936, 1349-2051, 1373-2070, 1441-2209, 1441-2389, 1442-1666, 1442-1737, 1442-1937, 1442-1967, 1442-1983, 1442-1999, 1442-2013, 1442-2036, 1442-2039, 1442-2059, 1442-2069, 1442-2072, 1442-2082, |
| | 1442-2088, 1442-2091, 1442-2109, 1442-2120, 1442-2201, 1442-2224, 1442-2346, 1442-2355, 1442-2356, 1443-2042, 1456-2188, 1480-2085, 1493-1748, 1499-1790, 1501-2003, 1505-2070, 1528-2178, 1551-1743, 1574-1845, 1582-2082, 1587-2567, |
| | 1590-2082, 1609-2298, 1620-2079, 1649-2515, 1680-2463, 1715-2444, 1718-2261, 1730-2410, 1751-2196, 1762-2812, 1780-2247, 1799-2336, 1799-2488, 1800-2636, 1824-2525, 1831-2685, 1855-2512, 1867-2682, 1870-2082, 1874-2708, 1883-2082, |
| | 1890-2082, 1899-2642, 1912-2463, 1925-2709, 2027-2746, 2028-2709, 2035-2726, 2050-2511, 2053-2574, 2060-2932, 2071-2569, 2074-2908, 2078-2913, 2102-2696, 2106-2830, 2115-2804, 2123-2793, 2144-2885, 2145-2851, 2146-2200, 2146-2315, |
| | 2146-2515, 2146-2555, 2146-2704, 2146-2873, 2161-2889, 2164-2819, 2167-2874, 2185-2749, 2186-3027, 2191-2425, 2206-2798, 2206-2850, 2224-2722, 2238-2814, 2253-2823, 2261-2819, 2267-2821, 2271-3076, 2273-3078, 2274-2961, 2275-2891, |
| | 2281-2883, 2282-3101, 2282-3250, 2290-3107, 2297-3143, 2310-2939, 2326-3140, 2329-3050, 2330-3090, 2330-3097, 2350-3015, 2361-3045, 2376-2794, 2378-3228, 2386-3090, 2394-3037, 2402-3020, 2416-3157, 2422-2973, 2422-3034, |

# Table 4

| Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length | Sequence Fragments |
|---|---|
| | 2430-3252, 2438-3096, 2443-3051, 2445-2902, 2445-3009, 2456-3311, 2460-3153, 2468-3061, 2474-3179, 2490-3061, 2497-3078, 2501-3409, 2503-3307, 2504-3234, 2506-3024, 2508-3250, 2511-3208, 2511-3337, 2512-3112, |
| | 2533-3144, 2533-3169, 2537-3223, 2537-3358, 2540-3156, 2541-3338, 2544-3069, 2544-3225, 2553-3222, 2553-3240, 2570-3086, 2591-3418, |
| | 2596-3129, 2600-3240, 2603-3316, 2608-3451, 2625-3141, 2627-3281, 2627-3425, 2630-3298, 2630-3459, 2631-3524, 2640-3178, 2641-3180, 2646-3157, |
| | 2650-3069, 2650-3458, 2653-3288, 2658-3230, 2660-3330, 2660-3496, 2667-3361, 2668-3070, 2668-3075, 2668-3101, 2668-3102, 2668-3193, 2668-3243, |
| | 2668-3252, 2668-3253, 2668-3260, 2668-3265, 2668-3277, 2668-3281, 2668-3297, 2668-3302, 2668-3337, 2668-3367, 2668-3369, 2668-3395, 2668-3411, |
| | 2668-3417, 2668-3511, 2668-3523, 2668-3529, 2668-3585, 2668-3590, 2668-3592, 2668-3597, 2668-3614, 2670-3558, 2671-3297, 2681-3522, 2682-3554, 2710-3083, |
| | 2712-3058, 2723-3606, 2724-3499, 2728-3547, 2730-3540, 2735-3674, 2741-3580, 2745-3521, 2759-3577, 2788-3572, 2811-3565, 2825-3730, 2828-3544, 2835-3619, |
| | 2843-3242, 2844-3532, 2844-3685, 2845-3545, 2845-3651, 2847-3554, 2848-3562, 2853-3731, 2856-3565, 2869-3598, 2871-3733, 2881-3601, 2908-3546, 2926-3578, |
| | 2927-3560, 2938-3553, 2942-3858, 2946-3824, 2947-3519, 2948-3867, 2958-3598, 2983-3499, 2996-3525, 2996-3526, 2997-3635, 2999-3797, 3003-3884, 3007-3691, |
| | 3013-3553, 3014-3813, 3035-3921, 3037-3497, 3046-3556, 3050-3698, 3051-3959, 3070-3813, 3071-3562, 3072-3752, 3078-3574, 3080-3522, 3088-3522, 3090-3521, 3099-3604, 3105-3833, 3111-3783, 3121-4035, 3127-3544, 3129-3613, 3129-3670, 3136-3845, 3141-3976, 3145-3761, 3145-3888, 3147-3574, 3153-4002, 3155-3774, 3156-4037, 3160-3591, 3161-3625, 3164-3888, 3169-3712, 3174-3712, 3176-3771, 3178-3733, 3181-3679, 3184-3866, 3190-3871, 3195-3766, 3195-3978, 3198-3994, 3199-3712, 3208-3683, 3208-3770, 3216-3509, 3218-3815, 3220-3688, 3223-3913, 3228-3647, 3228-3908, 3235-4154, 3239-3519, 3239-3772, 3239-3776, 3242-3761, 3245-3534, 3263-4110, 3267-3997, 3268-3823, 3274-3968, 3276-3841, 3285-3870, 3285-3998, 3292-3631, 3293-3812, 3294-3819, 3299-4019, 3302-3920, 3310-4213, 3327-3882, 3333-3611, 3335-3800, 3338-3748, 3339-4041, 3349-4012, 3361-3906, 3367-4213, 3374-4160, 3381-3763, 3381-4058, 3382-4061, 3382-4156, 3394-3490, 3399-4028, 3410-3835, 3417-4213, 3418-3702, 3422-4035, 3424-4018, 3426-4213, 3436-4004, 3441-3882, 3451-4154, 3460-3671, 3465-4095, 3470-4065, 3472-3596, 3473-3729, 3480-3730, 3480-3731, 3480-3741, 3486-4213, |
| | 3493-4133, 3502-3766, 3515-4074, 3517-4208, 3526-4031, 3542-4096, 3546-4131, 3946-3993, 3946-4033, 3958-4016, 3958-4032, 3968-3999, |
| | 3968-4033, 4007-4109, 4014-4056, 4014-4057, 4014-4068, 4014-4074, 4014-4087, 4014-4094, 4014-4097, 4014-4109, 4063-4109, 4063-4145, 4063-4149 |
| 22/7491216CB1/ 1784 | 1-788, 1-1784, 345-1179, 487-1167, 819-1005, 819-1007, 953-1167, 1130-1781 |

# Table 4

| Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length | Sequence Fragments |
|---|---|
| 23/71624817CB1/ 2141 | 1-639, 1-2125, 138-647, 138-737, 254-731, 366-490, 368-612, 386-2141, 612-1004, 699-737, 848-897, 848-1102 |
| 24/6945964CB1/ 2518 | 1-780, 102-917, 109-1154, 141-766, 142-673, 142-678, 142-679, 142-699, 142-702, 142-721, 142-725, 142-736, 142-755, 142-756, 142-761, 142-769, |
| | 142-772, 142-786, 142-790, 142-799, 142-828, 142-838, 142-853, 142-877, 142-881, 142-888, 142-889, 142-890, 142-977, 142-982, 142-987, 143-801, 146-678, |
| | 146-861, 146-914, 146-929, 146-1056, 160-814, 160-858, 166-827, 182-914, 183-713, 184-1095, 188-912, 211-777, 220-965, 233-1037, 235-1192, 236-785, |
| | 260-1028, 271-1196, 272-890, 282-1034, 308-1040, 319-842, 321-1047, 326-1089, 343-984, 348-1056, 397-957, 400-1223, 409-1139, 412-875, 418-1029, |
| | 422-1040, 433-898, 436-943, 438-800, 446-999, 470-1107, 472-978, 473-945, 483-1103, 510-773, 519-942, 523-674, 530-1028, 532-1084, 535-1028, |
| | 543-673, 545-1040, 563-1062, 572-998, 575-1026, 584-1077, 663-1064, 686-1104, 714-914, 714-1040, 714-1058, 714-1070, 714-1071, 728-975, |
| | 796-1035, 1051-1666, 1178-1841, 1410-1468, 1414-1444, 1416-1478, 1444-1498, 1450-1498, 1494-1968, 1537-1975, 1651-1980, 1908-2518 |

# Table 5

| Polynucleotide SEQ ID NO: | Incyte Project ID: | Representative Library |
|---|---|---|
| 13 | 2425607CB1 | BRAYDIN03 |
| 14 | 2786919CB1 | EPIPUNA01 |
| 15 | 1801130CB1 | TLYMNOT08 |
| 16 | 3535146CB1 | PITUDIR01 |
| 17 | 1436543CB1 | SINTFER02 |
| 18 | 7491063CB1 | BRAVTXT03 |
| 19 | 7625645CB1 | KIDNFEE02 |
| 20 | 5730123CB1 | ADRENOT11 |
| 21 | 7481031CB1 | BRAWTDR02 |
| 23 | 71624817CB1 | LIVRTUN04 |
| 24 | 6945964CB1 | THYRDIE01 |

# Table 6

| Library | Vector | Library Description |
|---|---|---|
| ADRENOT11 | pINCY | Library was constructed using RNA isolated from adrenal tissue removed from a 43-year-old Caucasian female during a unilateral adrenalectomy. Pathology for the associated tumor tissue indicated pheochromocytoma. |
| BRAVTXT03 | pINCY | The library was constructed using RNA isolated from treated astrocytes removed from the brain of a female fetus who died after 22 weeks'gestation. The cells were treated with tumor necrosis factor (TNF) alpha andinterleukin 1 (IL-1), 10ng/ml each for 24 hours. |
| BRAWTDR02 | PCDNA2.1 | This random primed library was constructed using RNA isolated from dentate nucleus tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver. |
| BRAYDIN03 | pINCY | This normalized library was constructed from 6.7 million independent clones from a brain tissue library. Starting RNA was made from RNA isolated from diseased hypothalamus tissue removed from a 57-year-old Caucasian male who died from a cerebrovascular accident. Patient history included Huntington's disease and emphysema. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 -hours/round) reannealing hybridization was used. The library was linearized and recircularized to select for insert containing clones. |
| EPIPUNA01 | PSPORT1 | Library was constructed using RNA isolated from untreated prostatic epithelial cell tissue removed from a 17-year-old Hispanic male. Serologies were negative. |
| KIDNFEE02 | PCDNA2.1 | This 5' biased random primed library was constructed using RNA isolated from kidney tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. Serology was negative. |
| LIVRTUN04 | pINCY | This normalized liver tumor cell line library was constructed from 1.72 million independent clones from a hepatocyte library. Starting RNA was isolated from an untreated C3A hepatocyte cell line, which is a derivative of a hepatoblastoma removed from a 15-year-old Caucasian male. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791,(except that a significantly longer (48 -hours/round) reannealing hybridization was used. |

# Table 6

| Library | Vector | Library Description |
|---|---|---|
| PITUDIR01 | PCDNA2.1 | This random primed library was constructed using RNA isolated from pituitary gland tissue removed from a 70-year-old female who died from metastatic adenocarcinoma. |
| SINTFER02 | pINCY | This random primed library was constructed using RNA isolated from small intestine tissue removed from a Caucasian male fetus who died from fetal demise. |
| THYRDIE01 | PCDNA2.1 | This 5' biased random primed library was constructed using RNA isolated from diseased thyroid tissue removed from a 22-year-old Caucasian female during closed thyroid biopsy, partial thyroidectomy, and regional lymph node excision. Pathology indicated adenomatous hyperplasia. The patient presented with malignant neoplasm of the thyroid. Patient history included normal delivery, alcohol abuse, and tobacco abuse. Previous surgeries included myringotomy. Patient medications included an unspecified type of birth control pills. Family history included hyperlipidemia and depressive disorder in the mother; and benign hypertension, congestive heart failure, and chronic leukemia in the grandparent(s). |
| TLYMNOT08 | pINCY | The library was constructed using RNA isolated from anergicallogenic T-lymphocyte tissue removed from an adult (40-50-year-old) Caucasian male.The cells were incubated for 3 days in the presence of 1 microgram/ml OKT3 mAb and 5% human serum. |

# Table 7

| Program | Description | Reference | Parameter Threshold |
|---|---|---|---|
| ABI FACTURA | A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences. | Applied Biosystems, Foster City, CA. | |
| ABI/PARACEL FDF | A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences. | Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA. | Mismatch <50% |
| ABI AutoAssembler | A program that assembles nucleic acid sequences. | Applied Biosystems, Foster City, CA. | |
| BLAST | A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx. | Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402. | ESTs: Probability value= 1.0E-8 or less; Full Length sequences: Probability value= 1.0E-10 or less |
| FASTA | A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch. | Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489. | ESTs: fasta E value=1.06E-6; Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less; Full Length sequences: fastx score=100 or greater |
| BLIMPS | A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions. | Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424. | Probability value= 1.0E-3 or less |

# Table 7

| Program | Description | Reference | Parameter Threshold |
|---|---|---|---|
| HMMER | An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART and TIGRFAM. | Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350. | PFAM, INCY, SMART, or TIGRFAM hits: Probability value= 1.0E-3 or less; Signal peptide hits: Score= 0 or greater |
| ProfileScan | An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite. | Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221. | Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1. |
| Phred | A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability. | Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194. | |
| Phrap | A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences. | Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA. | Score= 120 or greater; Match length= 56 or greater |
| Consed | A graphical tool for viewing and editing Phrap assemblies. | Gordon, D. et al. (1998) Genome Res. 8:195-202. | |
| SPScan | A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides. | Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439. | Score=3.5 or greater |
| TMAP | A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation. | Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371. | |

Table 7

| Program | Description | Reference | Parameter Threshold |
|---|---|---|---|
| TMHMMER | A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation. | Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182. | |
| Motifs | A program that searches amino acid sequences for patterns that matched those defined in Prosite. | Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI. | |

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-7 and SEQ ID NO:9-12,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, and

d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:13-24,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain

reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

19. A method for treating a disease or condition associated with decreased expression of functional ENZM, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional ENZM, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional

ENZM, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target

polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of ENZM in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of ENZM in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of ENZM in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim

11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-12, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-12.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:13.

69. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:14.

70. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:15.

71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.

72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.

73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.

74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.

75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID

NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

<110> INCYTE GENOMICS, INC.
TANG, Y. Tom
YUE, Henry
SANJANWALA, Madhusudan M.
RAMKUMAR, Jayalaxmi
YAO, Monique G.
SWARNAKAR, Anita
DING, Li
ELLIOTT, Vicki S.
GRIFFIN, Jennifer A.
LI, Joana X.
LAL, Preeti G.
LU, Dyung Aina M.
LU, Yan
GORVAD, Ann E.
FORSYTHE, Ian J.
DUGGAN, Brendan M.
THANGAVELU, Kavitha
EMERLING, Brooke M.
HAFALIA, April J.A.
BAUGHN, Mariah R.
BECHA, Shanya D.
SPRAGUE, William W.

<120> ENZYMES

<130> PI-0403 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/283,793; 60/291,544; 60/293,572; 60/308,182; 60/311,447;
60/315,874; 60/322,181
<151> 2001-04-13; 2001-05-16; 2001-05-25; 2001-07-27; 2001-08-09
2001-08-29; 2001-09-14

<160> 24

<170> PERL Program

<210> 1
<211> 363
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2425607CD1

<400> 1
Met Ala Arg Leu Arg Asp Cys Leu Pro Arg Leu Met Leu Thr Leu
1 5 10 15
Arg Ser Leu Leu Phe Trp Ser Leu Val Tyr Cys Tyr Cys Gly Leu
20 25 30
Cys Ala Ser Ile His Leu Leu Lys Leu Leu Trp Ser Leu Gly Lys
35 40 45
Gly Pro Arg Arg Pro Ser Gly Gly Pro Pro Gly Ser Thr Leu Pro
50 55 60
Ala Cys Leu Ser Asp Pro Ser Leu Gly Thr His Cys Tyr Val Arg
65 70 75
Ile Lys Asp Ser Gly Leu Arg Phe His Tyr Val Ala Ala Gly Glu
80 85 90
Arg Gly Lys Pro Leu Met Leu Leu Leu His Gly Phe Pro Glu Phe

```
                95                  100                 105
Trp Tyr Ile Ser Gly Gly Asn Gln Leu Arg Glu Phe Lys Ser Glu
                110                 115                 120
Tyr Arg Val Val Ala Leu Asp Leu Arg Gly Tyr Gly Glu Thr Asp
                125                 130                 135
Ala Pro Ile His Arg Gln Asn Tyr Lys Leu Asp Cys Leu Ile Thr
                140                 145                 150
Asp Ile Lys Asp Ile Leu Asp Ser Leu Gly Tyr Ser Lys Cys Val
                155                 160                 165
Leu Ile Gly His Asp Trp Gly Gly Met Ile Ala Trp Leu Ile Ala
                170                 175                 180
Ile Cys Tyr Pro Glu Met Val Met Lys Leu Ile Val Ile Asn Phe
                185                 190                 195
Pro His Pro Asn Val Phe Thr Glu Tyr Ile Leu Arg His Pro Ala
                200                 205                 210
Gln Leu Leu Lys Ser Ser Tyr Tyr Tyr Phe Phe Gln Ile Pro Trp
                215                 220                 225
Phe Pro Glu Phe Met Phe Ser Ile Asn Asp Phe Lys Val Leu Lys
                230                 235                 240
His Leu Phe Thr Ser His Ser Thr Gly Ile Gly Arg Lys Gly Cys
                245                 250                 255
Gln Leu Thr Thr Glu Asp Leu Glu Ala Tyr Ile Tyr Val Phe Ser
                260                 265                 270
Gln Pro Gly Ala Leu Ser Gly Pro Ile Asn His Tyr Arg Asn Ile
                275                 280                 285
Phe Ser Cys Leu Pro Leu Lys His His Met Val Thr Thr Pro Thr
                290                 295                 300
Leu Leu Leu Trp Gly Glu Asn Asp Ala Phe Met Glu Val Glu Met
                305                 310                 315
Ala Glu Val Thr Lys Ile Tyr Val Lys Asn Tyr Phe Arg Leu Thr
                320                 325                 330
Ile Leu Ser Glu Ala Ser His Trp Leu Gln Gln Asp Gln Pro Asp
                335                 340                 345
Ile Val Asn Lys Leu Ile Trp Thr Phe Leu Lys Glu Glu Thr Arg
                350                 355                 360
Lys Lys Asp


<210> 2
<211> 864
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2786919CD1

<400> 2
Met Pro Asn Val Leu Leu Pro Pro Lys Glu Ser Asn Leu Phe Lys
  1                 5                  10                  15
Arg Ile Leu Lys Cys Tyr Glu Gln Lys Gln Tyr Lys Asn Gly Leu
                20                  25                  30
Lys Phe Cys Lys Met Ile Leu Ser Asn Pro Lys Phe Ala Glu His
                35                  40                  45
Gly Glu Thr Leu Ala Met Lys Gly Leu Thr Leu Asn Cys Leu Gly
                50                  55                  60
Lys Lys Glu Glu Ala Tyr Glu Phe Val Arg Lys Gly Leu Arg Asn
                65                  70                  75
Asp Val Lys Ser His Val Cys Trp His Val Tyr Gly Leu Leu Gln
                80                  85                  90
Arg Ser Asp Lys Lys Tyr Asp Glu Ala Ile Lys Cys Tyr Arg Asn
                95                  100                 105
Ala Leu Lys Leu Asp Lys Asp Asn Leu Gln Ile Leu Arg Asp Leu
```

```
                110                 115                 120
Ser Leu Leu Gln Ile Gln Met Arg Asp Leu Glu Gly Tyr Arg Glu
                125                 130                 135
Thr Arg Tyr Gln Leu Leu Gln Leu Arg Pro Thr Gln Arg Ala Ser
                140                 145                 150
Trp Ile Gly Tyr Ala Ile Ala Tyr His Leu Leu Lys Asp Tyr Asp
                155                 160                 165
Met Ala Leu Lys Leu Leu Glu Glu Phe Arg Gln Thr Gln Gln Val
                170                 175                 180
Pro Pro Asn Lys Ile Asp Tyr Glu Tyr Ser Glu Leu Ile Leu Tyr
                185                 190                 195
Gln Asn Gln Val Met Arg Glu Ala Asp Leu Leu Gln Glu Ser Leu
                200                 205                 210
Glu His Ile Glu Met Tyr Glu Lys Gln Ile Cys Asp Lys Leu Leu
                215                 220                 225
Val Glu Glu Ile Lys Gly Glu Ile Leu Leu Lys Leu Gly Arg Leu
                230                 235                 240
Lys Glu Ala Ser Glu Val Phe Lys Asn Leu Ile Asp Arg Asn Ala
                245                 250                 255
Glu Asn Trp Cys Tyr Tyr Glu Gly Leu Glu Lys Ala Leu Gln Ile
                260                 265                 270
Ser Thr Leu Glu Glu Arg Leu Gln Ile Tyr Glu Glu Ile Ser Lys
                275                 280                 285
Gln His Pro Lys Ala Ile Thr Pro Arg Arg Leu Pro Leu Thr Leu
                290                 295                 300
Val Pro Gly Glu Arg Phe Arg Glu Leu Met Asp Lys Phe Leu Arg
                305                 310                 315
Val Asn Phe Ser Lys Gly Cys Pro Pro Leu Phe Thr Thr Leu Lys
                320                 325                 330
Ser Leu Tyr Tyr Asn Thr Glu Lys Val Ser Ile Ile Gln Glu Leu
                335                 340                 345
Val Thr Asn Tyr Glu Ala Ser Leu Lys Thr Cys Asp Phe Phe Ser
                350                 355                 360
Pro Tyr Glu Asn Gly Glu Lys Glu Pro Pro Thr Thr Leu Leu Trp
                365                 370                 375
Val Gln Tyr Phe Leu Ala Gln His Phe Asp Lys Leu Gly Gln Tyr
                380                 385                 390
Ser Leu Ala Leu Asp Tyr Ile Asn Ala Ala Ile Ala Ser Thr Pro
                395                 400                 405
Thr Leu Ile Glu Leu Phe Tyr Met Lys Ala Lys Ile Tyr Lys His
                410                 415                 420
Ile Gly Asn Leu Lys Glu Ala Ala Lys Trp Met Asp Glu Ala Gln
                425                 430                 435
Ser Leu Asp Thr Ala Asp Arg Phe Ile Asn Ser Lys Cys Ala Lys
                440                 445                 450
Tyr Met Leu Arg Ala Asn Met Ile Lys Glu Ala Glu Glu Met Cys
                455                 460                 465
Ser Lys Phe Thr Arg Glu Gly Thr Ser Ala Met Glu Asn Leu Asn
                470                 475                 480
Glu Met Gln Cys Met Trp Phe Gln Thr Glu Cys Ile Ser Ala Tyr
                485                 490                 495
Gln Arg Leu Gly Arg Tyr Gly Asp Ala Leu Lys Lys Cys His Glu
                500                 505                 510
Val Glu Arg His Phe Phe Glu Ile Thr Asp Asp Gln Phe Asp Phe
                515                 520                 525
His Thr Tyr Cys Met Arg Lys Met Thr Leu Arg Ala Tyr Val Asp
                530                 535                 540
Leu Leu Arg Leu Glu Asp Ile Leu Arg Arg His Ala Phe Tyr Phe
                545                 550                 555
Lys Ala Ala Arg Ser Ala Ile Glu Ile Tyr Leu Lys Leu Tyr Asp
                560                 565                 570
Asn Pro Leu Thr Asn Glu Ser Lys Gln Gln Glu Ile Asn Ser Glu
                575                 580                 585
```

```
Asn Leu Ser Ala Lys Glu Leu Lys Lys Met Leu Ser Lys Gln Arg
                590                 595                 600
Arg Ala Gln Lys Lys Ala Lys Leu Glu Glu Glu Arg Lys His Ala
                605                 610                 615
Glu Arg Glu Arg Gln Gln Lys Asn Gln Lys Lys Lys Arg Asp Glu
                620                 625                 630
Glu Glu Glu Glu Ala Ser Gly Leu Lys Glu Glu Leu Ile Pro Glu
                635                 640                 645
Lys Leu Glu Arg Val Glu Asn Pro Leu Glu Glu Ala Val Lys Phe
                650                 655                 660
Leu Ile Pro Leu Lys Asn Leu Val Ala Asp Asn Ile Asp Thr His
                665                 670                 675
Leu Leu Ala Phe Glu Ile Tyr Phe Arg Lys Gly Lys Phe Leu Leu
                680                 685                 690
Met Leu Gln Ser Val Lys Arg Ala Phe Ala Ile Asn Ser Asn Asn
                695                 700                 705
Pro Trp Leu His Glu Cys Leu Ile Arg Phe Ser Lys Ser Val Ser
                710                 715                 720
Asn His Ser Asn Leu Pro Asp Ile Val Ser Lys Val Leu Ser Gln
                725                 730                 735
Glu Met Gln Lys Ile Phe Val Lys Lys Asp Leu Glu Ser Phe Asn
                740                 745                 750
Glu Asp Phe Leu Lys Arg Asn Ala Thr Ser Leu Gln His Leu Leu
                755                 760                 765
Ser Gly Ala Lys Met Met Tyr Phe Leu Asp Lys Ser Arg Gln Glu
                770                 775                 780
Lys Ala Ile Ala Ile Ala Thr Arg Leu Asp Glu Thr Ile Lys Asp
                785                 790                 795
Lys Asp Val Lys Thr Leu Ile Lys Val Ser Glu Ala Leu Leu Asp
                800                 805                 810
Gly Ser Phe Gly Asn Cys Ser Ser Gln Tyr Glu Glu Tyr Arg Met
                815                 820                 825
Ala Cys His Asn Leu Leu Pro Phe Thr Ser Ala Phe Leu Pro Ala
                830                 835                 840
Val Asn Glu Val Asp Asn Pro Asn Val Ala Leu Asn His Thr Ala
                845                 850                 855
Asn Tyr Asp Val Leu Ala Asn Glu Ile
                860

<210> 3
<211> 376
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1801130CD1

<400> 3
Met Ser Ala Gln Val His Arg Gln Lys Gly Leu Asp Leu Ser Gln
  1               5                  10                  15
Ile Pro Tyr Phe Asn Leu Val Lys His Leu Thr Pro Ala Cys Pro
                20                  25                  30
Asn Val Tyr Ser Ile Ser Gln Phe His His Thr Thr Pro Asp Ser
                35                  40                  45
Lys Thr His Ser Gly Glu Lys Tyr Thr Asp Pro Phe Lys Leu Gly
                50                  55                  60
Trp Arg Asp Leu Lys Gly Leu Tyr Glu Asp Ile Arg Lys Glu Leu
                65                  70                  75
Leu Ile Ser Thr Ser Glu Leu Lys Glu Met Ser Glu Tyr Tyr Phe
                80                  85                  90
Asp Gly Lys Gly Lys Ala Phe Arg Pro Ile Ile Val Ala Leu Met
                95                 100                 105
```

```
Ala Arg Ala Cys Asn Ile His His Asn Asn Ser Arg His Val Gln
                110                 115                 120
Ala Ser Gln Arg Ala Ile Ala Leu Ile Ala Glu Met Ile His Thr
                125                 130                 135
Ala Ser Leu Val His Asp Asp Val Ile Asp Asp Ala Ser Ser Arg
                140                 145                 150
Arg Gly Lys His Thr Val Asn Lys Ile Trp Gly Glu Lys Lys Ala
                155                 160                 165
Val Leu Ala Gly Asp Leu Ile Leu Ser Ala Ala Ser Ile Ala Leu
                170                 175                 180
Ala Arg Ile Gly Asn Thr Thr Val Ile Ser Ile Leu Thr Gln Val
                185                 190                 195
Ile Glu Asp Leu Val Arg Gly Glu Phe Leu Gln Leu Gly Ser Lys
                200                 205                 210
Glu Asn Glu Asn Glu Arg Phe Ala His Tyr Leu Glu Lys Thr Phe
                215                 220                 225
Lys Lys Thr Ala Ser Leu Ile Ala Asn Ser Cys Lys Ala Val Ser
                230                 235                 240
Val Leu Gly Cys Pro Asp Pro Val Val His Glu Ile Ala Tyr Gln
                245                 250                 255
Tyr Gly Lys Asn Val Gly Ile Ala Phe Gln Leu Ile Asp Asp Val
                260                 265                 270
Leu Asp Phe Thr Ser Cys Ser Asp Gln Met Gly Lys Pro Thr Ser
                275                 280                 285
Ala Asp Leu Lys Leu Gly Leu Ala Thr Gly Pro Val Leu Phe Ala
                290                 295                 300
Cys Gln Gln Phe Pro Glu Met Asn Ala Met Ile Met Arg Arg Phe
                305                 310                 315
Ser Leu Pro Gly Asp Val Asp Arg Ala Arg Gln Tyr Val Leu Gln
                320                 325                 330
Ser Asp Gly Val Gln Gln Thr Thr Tyr Leu Ala Gln Gln Tyr Cys
                335                 340                 345
His Glu Ala Ile Arg Glu Ile Ser Lys Leu Arg Pro Ser Pro Glu
                350                 355                 360
Arg Asp Ala Leu Ile Gln Leu Ser Glu Ile Val Leu Thr Arg Asp
                365                 370                 375
Lys
```

<210> 4
<211> 978
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3535146CD1

<400> 4

```
Met Gly Thr Arg Leu Pro Leu Val Leu Arg Gln Leu Arg Arg Pro
  1               5                  10                  15
Pro Gln Pro Pro Gly Pro Pro Arg Arg Leu Arg Val Pro Cys Arg
                20                  25                  30
Ala Ser Ser Gly Gly Gly Gly Gly Gly Gly Gly Gly Arg Glu Gly
                35                  40                  45
Leu Leu Gly Gln Arg Arg Pro Gln Asp Gly Gln Ala Arg Ser Ser
                50                  55                  60
Cys Ser Pro Gly Gly Arg Thr Pro Ala Ala Arg Asp Ser Ile Val
                65                  70                  75
Arg Glu Val Ile Gln Asn Ser Lys Glu Val Leu Ser Leu Leu Gln
                80                  85                  90
Glu Lys Asn Pro Ala Phe Lys Pro Val Leu Ala Ile Ile Gln Ala
                95                 100                 105
```

```
Gly Asp Asp Asn Leu Met Gln Glu Ile Asn Gln Asn Leu Ala Glu
                110                 115                 120
Glu Ala Gly Leu Asn Ile Thr His Ile Cys Leu Pro Pro Asp Ser
                125                 130                 135
Ser Glu Ala Glu Ile Ile Asp Glu Ile Leu Lys Ile Asn Glu Asp
                140                 145                 150
Thr Arg Val His Gly Leu Ala Leu Gln Ile Ser Glu Asn Leu Phe
                155                 160                 165
Ser Asn Lys Val Leu Asn Ala Leu Lys Pro Glu Lys Asp Val Asp
                170                 175                 180
Gly Val Thr Asp Ile Asn Leu Gly Lys Leu Val Arg Gly Asp Ala
                185                 190                 195
His Glu Cys Phe Val Ser Pro Val Ala Lys Ala Val Ile Glu Leu
                200                 205                 210
Leu Glu Lys Ser Gly Val Asn Leu Asp Gly Lys Lys Ile Leu Val
                215                 220                 225
Val Gly Ala His Gly Ser Leu Glu Ala Ala Leu Gln Cys Leu Phe
                230                 235                 240
Gln Arg Lys Gly Ser Met Thr Met Ser Ile Gln Trp Lys Thr Arg
                245                 250                 255
Gln Leu Gln Ser Lys Leu His Glu Ala Asp Ile Val Val Leu Gly
                260                 265                 270
Ser Pro Lys Pro Glu Glu Ile Pro Leu Thr Trp Ile Gln Pro Gly
                275                 280                 285
Thr Thr Val Leu Asn Cys Ser His Asp Phe Leu Ser Gly Lys Val
                290                 295                 300
Gly Cys Gly Ser Pro Arg Ile His Phe Gly Gly Leu Ile Glu Glu
                305                 310                 315
Asp Asp Val Ile Leu Leu Ala Ala Ala Leu Arg Ile Gln Asn Met
                320                 325                 330
Val Ser Ser Gly Arg Arg Trp Leu Arg Glu Gln Gln His Arg Arg
                335                 340                 345
Trp Arg Leu His Cys Leu Lys Leu Gln Pro Leu Ser Pro Val Pro
                350                 355                 360
Ser Asp Ile Glu Ile Ser Arg Gly Gln Thr Pro Lys Ala Val Asp
                365                 370                 375
Val Leu Ala Lys Glu Ile Gly Leu Leu Ala Asp Glu Ile Glu Ile
                380                 385                 390
Tyr Gly Lys Ser Lys Ala Lys Val Arg Leu Ser Val Leu Glu Arg
                395                 400                 405
Leu Lys Asp Gln Ala Asp Gly Lys Tyr Val Leu Val Ala Gly Ile
                410                 415                 420
Thr Pro Thr Pro Leu Gly Glu Gly Lys Ser Thr Val Thr Ile Gly
                425                 430                 435
Leu Val Gln Ala Leu Thr Ala His Leu Asn Val Asn Ser Phe Ala
                440                 445                 450
Cys Leu Arg Gln Pro Ser Gln Gly Pro Thr Phe Gly Val Lys Gly
                455                 460                 465
Gly Ala Ala Gly Gly Gly Tyr Ala Gln Val Ile Pro Met Glu Glu
                470                 475                 480
Phe Asn Leu His Leu Thr Gly Asp Ile His Ala Ile Thr Ala Ala
                485                 490                 495
Asn Asn Leu Leu Ala Ala Ala Ile Asp Thr Arg Ile Leu His Glu
                500                 505                 510
Asn Thr Gln Thr Asp Lys Ala Leu Tyr Asn Arg Leu Val Pro Leu
                515                 520                 525
Val Asn Gly Val Arg Glu Phe Ser Glu Ile Gln Leu Ala Arg Leu
                530                 535                 540
Lys Lys Leu Gly Ile Asn Lys Thr Asp Pro Ser Thr Leu Thr Glu
                545                 550                 555
Glu Glu Val Ser Lys Phe Ala Arg Leu Asp Ile Asp Pro Ser Thr
                560                 565                 570
Ile Thr Trp Gln Arg Val Leu Asp Thr Asn Asp Arg Phe Leu Arg
```

575 580 585
Lys Ile Thr Ile Gly Gln Gly Asn Thr Glu Lys Gly His Tyr Arg
590 595 600
Gln Ala Gln Phe Asp Ile Ala Val Ala Ser Glu Ile Met Ala Val
605 610 615
Leu Ala Leu Thr Asp Ser Leu Ala Asp Met Lys Ala Arg Leu Gly
620 625 630
Arg Met Val Val Ala Ser Asp Lys Ser Gly Gln Pro Val Thr Ala
635 640 645
Asp Asp Leu Gly Val Thr Gly Ala Leu Thr Val Leu Met Lys Asp
650 655 660
Ala Ile Lys Pro Asn Leu Met Gln Thr Leu Glu Gly Thr Pro Val
665 670 675
Phe Val His Ala Gly Pro Phe Ala Asn Ile Ala His Gly Asn Ser
680 685 690
Ser Val Leu Ala Asp Lys Ile Ala Leu Lys Leu Val Gly Glu Glu
695 700 705
Gly Phe Val Val Thr Glu Ala Gly Phe Gly Ala Asp Ile Gly Met
710 715 720
Glu Lys Phe Phe Asn Ile Lys Cys Arg Ala Ser Gly Leu Val Pro
725 730 735
Asn Val Val Val Leu Val Ala Thr Val Arg Ala Leu Lys Met His
740 745 750
Gly Gly Gly Pro Ser Val Thr Ala Gly Val Pro Leu Lys Lys Glu
755 760 765
Tyr Thr Glu Glu Asn Ile Gln Leu Val Ala Asp Gly Cys Cys Asn
770 775 780
Leu Gln Lys Gln Ile Gln Ile Thr Gln Leu Phe Gly Val Pro Val
785 790 795
Val Val Ala Leu Asn Val Phe Lys Thr Asp Thr Arg Ala Glu Ile
800 805 810
Asp Leu Val Cys Glu Leu Ala Lys Arg Ala Gly Ala Phe Asp Ala
815 820 825
Val Pro Cys Tyr His Trp Ser Val Gly Gly Lys Gly Ser Val Asp
830 835 840
Leu Ala Arg Ala Val Arg Glu Ala Ala Ser Lys Arg Ser Arg Phe
845 850 855
Gln Phe Leu Tyr Asp Val Gln Val Pro Ile Val Asp Lys Ile Arg
860 865 870
Thr Ile Ala Gln Ala Val Tyr Gly Ala Lys Asp Ile Glu Leu Ser
875 880 885
Pro Glu Ala Gln Ala Lys Ile Asp Arg Tyr Thr Gln Gln Gly Phe
890 895 900
Gly Asn Leu Pro Ile Cys Met Ala Lys Thr His Leu Ser Leu Ser
905 910 915
His Gln Pro Asp Lys Lys Gly Val Pro Arg Asp Phe Ile Leu Pro
920 925 930
Ile Ser Asp Val Arg Ala Ser Ile Gly Ala Gly Phe Ile Tyr Pro
935 940 945
Leu Val Gly Thr Met Ser Thr Met Pro Gly Leu Pro Thr Arg Pro
950 955 960
Cys Phe Tyr Asp Ile Asp Leu Asp Thr Glu Thr Glu Gln Val Lys
965 970 975
Gly Leu Phe

<210> 5
<211> 349
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 1436543CD1

<400> 5
Met Ala Ala Ser Glu Ala Ala Val Val Ser Ser Pro Ser Leu Lys
1 5 10 15
Thr Asp Thr Ser Pro Val Leu Glu Thr Ala Gly Thr Val Ala Ala
20 25 30
Met Ala Ala Thr Pro Ser Ala Arg Ala Ala Ala Ala Val Val Ala
35 40 45
Ala Ala Ala Arg Thr Gly Ser Glu Ala Arg Val Ser Lys Ala Ala
50 55 60
Leu Ala Thr Lys Leu Leu Ser Leu Ser Gly Val Phe Ala Val His
65 70 75
Lys Pro Lys Gly Pro Thr Ser Ala Glu Leu Leu Asn Arg Leu Lys
80 85 90
Glu Lys Leu Leu Ala Glu Ala Gly Met Pro Ser Pro Glu Trp Thr
95 100 105
Lys Arg Lys Lys Gln Thr Leu Lys Ile Gly His Gly Gly Thr Leu
110 115 120
Asp Ser Ala Ala Arg Gly Val Leu Val Val Gly Ile Gly Ser Gly
125 130 135
Thr Lys Met Leu Thr Ser Met Leu Ser Gly Ser Lys Arg Tyr Thr
140 145 150
Ala Ile Gly Glu Leu Gly Lys Ala Thr Asp Thr Leu Asp Ser Thr
155 160 165
Gly Arg Val Thr Glu Glu Lys Pro Tyr Asp Lys Ile Thr Gln Glu
170 175 180
Asp Ile Glu Gly Ile Leu Gln Lys Phe Thr Gly Asn Ile Met Gln
185 190 195
Val Pro Pro Leu Tyr Ser Ala Leu Lys Lys Asp Gly Gln Arg Leu
200 205 210
Ser Thr Leu Met Lys Arg Gly Glu Val Val Glu Ala Lys Pro Ala
215 220 225
Arg Pro Val Thr Val Tyr Ser Ile Ser Leu Gln Lys Phe Gln Pro
230 235 240
Pro Phe Phe Thr Leu Asp Val Glu Cys Gly Gly Gly Phe Tyr Ile
245 250 255
Arg Ser Leu Val Ser Asp Ile Gly Lys Glu Leu Ser Ser Cys Ala
260 265 270
Asn Val Leu Glu Leu Thr Arg Thr Lys Gln Gly Pro Phe Thr Leu
275 280 285
Glu Glu His Ala Leu Pro Glu Asp Lys Trp Thr Ile Asp Asp Ile
290 295 300
Ala Gln Ser Leu Glu His Cys Ser Ser Leu Phe Pro Ala Glu Leu
305 310 315
Ala Leu Lys Lys Ser Lys Pro Glu Ser Asn Glu Gln Val Leu Ser
320 325 330
Cys Glu Tyr Ile Thr Leu Asn Glu Pro Lys Arg Glu Asp Asp Val
335 340 345
Ile Lys Thr Cys

<210> 6
<211> 399
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7491063CD1

<400> 6
Met Phe Ser Lys Gly Ser Met Val Leu Ala Tyr Ser Ala Gly Leu

```
  1               5                  10                  15
Asp Thr Ser Cys Ile Leu Val Trp Leu Lys Glu Gln Gly Tyr Asp
             20                  25                  30
Ile Ile Ala Tyr Leu Ala Asn Val Gly Gln Lys Glu Asp Phe Lys
             35                  40                  45
Glu Ala Arg Lys Lys Ala Leu Asn Leu Gly Ser Lys Lys Val Phe
             50                  55                  60
Ile Glu Asp Val Ser Lys Glu Phe Val Glu Glu Phe Ile Trp Pro
             65                  70                  75
Tyr Glu Asp Cys Tyr Leu Leu Gly Pro Ser Leu Pro Arg Pro Tyr
             80                  85                  90
Ile Thr Arg Lys Gln Val Glu Ile Ala Gln Trp Glu Gly Ala Lys
             95                 100                 105
Tyr Val Ser His Ser Ala Met Gly Lys Gly Asn Asp Gln Val Trp
            110                 115                 120
Phe Glu Leu Ala Cys Tyr Ser Leu Ala Pro Gln Ile Lys Val Ile
            125                 130                 135
Ala Pro Gly Arg Ile Pro Glu Phe Tyr Asn Gln Ser Lys Gly Arg
            140                 145                 150
Ser Asp Leu Met Glu Tyr Ala Glu Lys His Gly Ile Pro Ile Pro
            155                 160                 165
Val Thr Leu Lys His Pro Trp Asn Met Asp Glu Asn Leu Met His
            170                 175                 180
Ile Ser His Glu Gly Trp Asn Leu Gly Glu Pro Gln Glu Pro Glu
            185                 190                 195
Ala Pro Ser Gly Leu Tyr Met Lys Ile Gln Asp Leu Ala Lys Ala
            200                 205                 210
Pro Asn Thr Pro Asn Ile Phe Lys Thr Glu Gly Val Pro Val Lys
            215                 220                 225
Val Thr Ser Ile Lys Asp Gly Thr Thr His Gln Thr Ser Leu Val
            230                 235                 240
Leu Leu His Val Thr Trp Asn Glu Val Ala Gly Lys His Ser Val
            245                 250                 255
Gly His Ile Asp Ile Val Glu Asn Arg Phe Ile Glu Met Asn Ile
            260                 265                 270
Cys Lys Thr Pro Ala Gly Thr Ile Leu Tyr His Pro His Leu Asp
            275                 280                 285
Ile Glu Gly Phe Ala Met Glu Gln Glu Val Arg Lys Ile Lys Gln
            290                 295                 300
Gly Leu Gly Leu Lys Phe Ala Glu Leu Val Tyr Thr Gly Phe Trp
            305                 310                 315
His Asn Pro Gln Cys Asp Phe Ala His His Cys Ile Ala Lys Ser
            320                 325                 330
Gln Asp Arg Val Glu Gly Lys Val Gln Val Ser Ile Phe Lys Gly
            335                 340                 345
Gln Val Tyr Ile Leu Cys Gln Glu Pro Pro Leu Ser Leu Tyr Ser
            350                 355                 360
Gly Glu Gln Val Ser Met Asn Val Glu Gly Asn Asp Glu Pro Ala
            365                 370                 375
Ser Arg Leu Ile Asn Ile Asn Ser Leu Arg Met Lys Glu Tyr His
            380                 385                 390
His Leu Gln Ser Lys Val Thr Ala Lys
            395
```

<210> 7
<211> 278
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7625645CD1

<400> 7
Met Glu Ile Ile Arg Thr Ile Ala Glu Leu Arg Val Arg Leu Ala
1 5 10 15
Arg Glu Ser Asn Ile Ala Leu Val Pro Thr Met Gly Asn Ile His
20 25 30
Asp Gly His Leu Ser Leu Val Arg Ile Gly Leu Lys Arg Ala Pro
35 40 45
Phe Thr Val Thr Ser Ile Phe Val Asn Arg Leu Gln Phe Ile Gln
50 55 60
Gly Glu Asp Phe Asp Lys Tyr Pro Arg Thr Phe Glu Glu Asp Cys
65 70 75
Glu Lys Leu Glu Ala Ala Gly Asn Ser Val Val Phe Ala Pro Asp
80 85 90
Glu Thr Gln Met Tyr Pro Glu Arg Gln Val Phe Ile Val Glu Pro
95 100 105
Pro Pro Ile Ala Asn Lys Leu Glu Gly Arg Phe Arg Pro Gly His
110 115 120
Phe Arg Gly Val Ser Thr Val Val Leu Lys Leu Phe Asn Met Val
125 130 135
Gln Pro Gln Val Ala Val Phe Gly Lys Lys Asp Tyr Gln Gln Leu
140 145 150
Ser Ile Val Arg His Met Val Asp Gln Leu Ala Leu Pro Leu Ser
155 160 165
Ile Ile Pro Ala Glu Thr Val Arg Ala Glu Asp Gly Leu Ala Leu
170 175 180
Ser Ser Arg Asn Arg Tyr Leu Ser Pro Glu Glu Arg Ser Glu Ala
185 190 195
Pro Arg Leu Asn Glu Ala Leu Arg Gln Val Lys Gln Ala Val Glu
200 205 210
Ser Gly Asp Arg Asn Phe Glu Ala Leu Glu Phe Ala Ala Asp Ala
215 220 225
Leu Leu Ala Arg His Arg Trp His Val Asp Tyr Val Val Val Arg
230 235 240
Ser Arg Arg Thr Leu Met Gln Pro Asp Pro Asp Glu Arg Glu Leu
245 250 255
Val Val Leu Gly Ala Ala Arg Leu Gly Gln Thr Arg Leu Ile Asp
260 265 270
Asn Leu Glu Ile Leu Ala Pro Ala
275

<210> 8
<211> 583
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5730123CD1

<400> 8
Met Ser Gly Val Arg Gly Leu Ser Arg Leu Leu Ser Ala Arg Arg
1 5 10 15
Leu Ala Leu Ala Lys Ala Trp Pro Thr Val Leu Gln Thr Gly Thr
20 25 30
Arg Gly Phe His Phe Thr Val Asp Gly Asn Lys Arg Ala Ser Ala
35 40 45
Lys Val Ser Asp Ser Ile Ser Ala Gln Tyr Pro Val Val Asp His
50 55 60
Glu Phe Asp Ala Val Val Val Gly Ala Gly Gly Ala Gly Leu Arg
65 70 75
Ala Ala Phe Gly Leu Ser Glu Ala Gly Phe Asn Thr Ala Cys Val
80 85 90
Thr Lys Leu Phe Pro Thr Arg Ser His Thr Val Ala Ala Gln Gly

```
                95                  100                 105
Gly Ile Asn Ala Ala Leu Gly Asn Met Glu Glu Asp Asn Trp Arg
               110                 115                 120
Trp His Phe Tyr Asp Thr Val Lys Gly Ser Asp Trp Leu Gly Asp
               125                 130                 135
Gln Asp Ala Ile His Tyr Met Thr Glu Gln Ala Pro Ala Ala Val
               140                 145                 150
Val Glu Leu Glu Asn Tyr Gly Met Pro Phe Ser Arg Thr Glu Asp
               155                 160                 165
Gly Lys Ile Tyr Gln Arg Ala Phe Gly Gly Gln Ser Leu Lys Phe
               170                 175                 180
Gly Lys Gly Gly Gln Ala His Arg Cys Cys Cys Val Ala Asp Arg
               185                 190                 195
Thr Gly His Ser Leu Leu His Thr Leu Tyr Gly Arg Ser Leu Arg
               200                 205                 210
Tyr Asp Thr Ser Tyr Phe Val Glu Tyr Phe Ala Leu Asp Leu Leu
               215                 220                 225
Met Glu Asn Gly Glu Cys Arg Gly Val Ile Ala Leu Cys Ile Glu
               230                 235                 240
Asp Gly Ser Ile His Arg Ile Arg Ala Lys Asn Thr Val Val Ala
               245                 250                 255
Thr Gly Gly Tyr Gly Arg Thr Tyr Phe Ser Cys Thr Ser Ala His
               260                 265                 270
Thr Ser Thr Gly Asp Gly Thr Ala Met Ile Thr Arg Ala Gly Leu
               275                 280                 285
Pro Cys Gln Asp Leu Glu Phe Val Gln Phe His Pro Thr Gly Ile
               290                 295                 300
Tyr Gly Ala Gly Cys Leu Ile Thr Glu Gly Cys Arg Gly Glu Gly
               305                 310                 315
Gly Ile Leu Ile Asn Ser Gln Gly Glu Arg Phe Met Glu Arg Tyr
               320                 325                 330
Ala Pro Val Ala Lys Asp Leu Ala Ser Arg Asp Val Val Ser Arg
               335                 340                 345
Ser Met Thr Leu Glu Ile Arg Glu Gly Arg Gly Cys Gly Pro Glu
               350                 355                 360
Lys Asp His Val Tyr Leu Gln Leu His His Leu Pro Pro Glu Gln
               365                 370                 375
Leu Ala Thr Arg Leu Pro Gly Ile Ser Glu Thr Ala Met Ile Phe
               380                 385                 390
Ala Gly Val Asp Val Thr Lys Glu Pro Ile Pro Val Leu Pro Thr
               395                 400                 405
Val His Tyr Asn Met Gly Gly Ile Pro Thr Asn Tyr Lys Gly Gln
               410                 415                 420
Val Leu Arg His Val Asn Gly Gln Asp Gln Ile Val Pro Gly Leu
               425                 430                 435
Tyr Ala Cys Gly Glu Ala Ala Cys Ala Ser Val His Gly Ala Asn
               440                 445                 450
Arg Leu Gly Ala Asn Ser Leu Leu Asp Leu Val Val Phe Gly Arg
               455                 460                 465
Ala Cys Ala Leu Ser Ile Glu Glu Ser Cys Arg Pro Gly Asp Lys
               470                 475                 480
Val Pro Pro Ile Lys Pro Asn Ala Gly Glu Glu Ser Val Met Asn
               485                 490                 495
Leu Asp Lys Leu Arg Phe Ala Asp Gly Ser Ile Arg Thr Ser Glu
               500                 505                 510
Leu Arg Leu Ser Met Gln Lys Val Arg Ile Asp Glu Tyr Asp Tyr
               515                 520                 525
Ser Lys Pro Ile Gln Gly Gln Gln Lys Lys Pro Phe Glu Glu His
               530                 535                 540
Trp Arg Lys His Thr Leu Ser Phe Val Asp Val Gly Thr Gly Lys
               545                 550                 555
Val Thr Leu Glu Tyr Arg Pro Val Ile Asp Lys Thr Leu Asn Glu
               560                 565                 570
```

```
Ala Asp Cys Ala Thr Ile Pro Pro Ala Ile Arg Ser Tyr
                575                 580

<210> 9
<211> 1032
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481031CD1

<400> 9
Met Gln Arg Pro Gly Gly Arg Trp Ala Gly Pro Glu Arg Val Gln
  1               5                  10                  15
Val Arg Arg Ala Leu Trp Gly Pro Cys Ser Leu Arg Gly Pro Glu
                20                  25                  30
Trp Ala Ala Pro Asp Thr Arg Glu Cys Val Gln Thr Arg Leu Leu
                35                  40                  45
Thr Ala Ala Ser Pro Ala Asp His Gly Gly Gly Arg Pro Ala Ala
                50                  55                  60
Ala Gly Pro Leu Leu Lys Arg Ser His Ser Val Pro Ala Pro Ser
                65                  70                  75
Ile Arg Lys Gln Ile Leu Glu Glu Leu Glu Lys Pro Gly Ala Gly
                80                  85                  90
Asp Pro Asp Pro Ser Ala Pro Gln Gly Gly Pro Gly Ser Ala Thr
                95                 100                 105
Ser Asp Leu Gly Met Ala Cys Asp Arg Ala Arg Val Phe Leu His
               110                 115                 120
Ser Asp Glu His Pro Gly Ser Ser Val Ala Ser Lys Ser Arg Lys
               125                 130                 135
Ser Val Met Val Ala Glu Ile Pro Ser Thr Val Ser Gln His Ser
               140                 145                 150
Glu Ser His Thr Asp Glu Thr Leu Ala Ser Arg Lys Ser Asp Ala
               155                 160                 165
Ile Phe Arg Ala Ala Lys Lys Asp Leu Leu Thr Leu Met Lys Leu
               170                 175                 180
Glu Asp Ser Ser Leu Leu Asp Gly Arg Val Ala Leu Leu His Val
               185                 190                 195
Pro Ala Gly Thr Val Val Ser Arg Gln Gly Asp Gln Asp Ala Ser
               200                 205                 210
Ile Leu Phe Val Val Ser Gly Leu Leu His Val Tyr Gln Arg Lys
               215                 220                 225
Ile Gly Ser Gln Glu Asp Thr Cys Leu Phe Leu Thr Arg Pro Gly
               230                 235                 240
Glu Met Val Gly Gln Leu Ala Val Leu Thr Gly Glu Pro Leu Ile
               245                 250                 255
Phe Thr Val Lys Ala Asn Arg Asp Cys Ser Phe Leu Ser Ile Ser
               260                 265                 270
Lys Ala His Phe Tyr Glu Ile Met Arg Lys Gln Pro Thr Val Val
               275                 280                 285
Leu Gly Val Ala His Thr Val Val Lys Arg Met Ser Ser Phe Val
               290                 295                 300
Arg Gln Ile Asp Phe Ala Leu Asp Trp Val Glu Val Glu Ala Gly
               305                 310                 315
Arg Ala Ile Tyr Arg Gln Gly Asp Lys Ser Asp Cys Thr Tyr Ile
               320                 325                 330
Met Leu Ser Gly Arg Leu Arg Ser Val Ile Arg Lys Asp Asp Gly
               335                 340                 345
Lys Lys Arg Leu Ala Gly Glu Tyr Gly Arg Gly Asp Leu Val Gly
               350                 355                 360
Val Val Glu Thr Leu Thr His Gln Ala Arg Ala Thr Thr Val His
               365                 370                 375
```

```
Ala Val Arg Asp Ser Glu Leu Ala Lys Leu Pro Ala Gly Ala Leu
                380                 385                 390
Thr Ser Ile Lys Arg Arg Tyr Pro Gln Val Val Thr Arg Leu Ile
                395                 400                 405
His Leu Leu Gly Glu Lys Ile Leu Gly Ser Leu Gln Gln Gly Pro
                410                 415                 420
Val Thr Gly His Gln Leu Gly Leu Pro Thr Glu Gly Ser Lys Trp
                425                 430                 435
Asp Leu Gly Asn Pro Ala Val Asn Leu Ser Thr Val Ala Val Met
                440                 445                 450
Pro Val Ser Glu Glu Val Pro Leu Thr Ala Phe Ala Leu Glu Leu
                455                 460                 465
Glu His Ala Leu Ser Ala Ile Gly Pro Thr Leu Leu Leu Thr Ser
                470                 475                 480
Asp Asn Ile Lys Arg Arg Leu Gly Ser Ala Ala Leu Asp Ser Val
                485                 490                 495
His Glu Tyr Arg Leu Ser Ser Trp Leu Gly Gln Gln Glu Asp Thr
                500                 505                 510
His Arg Ile Val Leu Tyr Gln Ala Asp Gly Thr Leu Thr Pro Trp
                515                 520                 525
Thr Gln Arg Cys Val Arg Gln Ala Asp Cys Ile Leu Ile Val Gly
                530                 535                 540
Leu Gly Asp Gln Glu Pro Thr Val Gly Glu Leu Glu Arg Met Leu
                545                 550                 555
Glu Ser Thr Ala Val Arg Ala Gln Lys Gln Leu Ile Leu Leu His
                560                 565                 570
Arg Glu Glu Gly Pro Ala Pro Ala Arg Thr Val Glu Trp Leu Asn
                575                 580                 585
Met Arg Ser Trp Cys Ser Gly His Leu His Leu Cys Cys Pro Arg
                590                 595                 600
Arg Val Phe Ser Arg Arg Ser Leu Pro Lys Leu Val Glu Met Tyr
                605                 610                 615
Lys His Val Phe Gln Arg Pro Pro Asp Arg His Ser Asp Phe Ser
                620                 625                 630
Arg Leu Ala Arg Val Leu Thr Gly Asn Ala Ile Ala Leu Val Leu
                635                 640                 645
Gly Gly Gly Gly Ala Arg Gly Cys Ala Gln Val Gly Val Leu Lys
                650                 655                 660
Ala Leu Ala Glu Cys Gly Ile Pro Val Asp Met Val Gly Gly Thr
                665                 670                 675
Ser Ile Gly Ala Phe Val Gly Ala Leu Tyr Ser Glu Glu Arg Asn
                680                 685                 690
Tyr Ser Gln Met Arg Ile Arg Ala Lys Gln Trp Ala Glu Gly Met
                695                 700                 705
Thr Ser Leu Met Lys Ala Ala Leu Asp Leu Thr Tyr Pro Ile Thr
                710                 715                 720
Ser Met Phe Ser Gly Ala Gly Phe Asn Ser Ser Ile Phe Ser Val
                725                 730                 735
Phe Lys Asp Gln Gln Ile Glu Asp Leu Trp Ile Pro Tyr Phe Ala
                740                 745                 750
Ile Thr Thr Asp Ile Thr Ala Ser Ala Met Arg Val His Thr Asp
                755                 760                 765
Gly Ser Leu Trp Arg Tyr Val Arg Ala Ser Met Ser Leu Ser Gly
                770                 775                 780
Tyr Met Pro Pro Leu Cys Asp Pro Lys Asp Gly His Leu Leu Met
                785                 790                 795
Asp Gly Gly Tyr Ile Asn Asn Leu Pro Ala Asp Val Ala Arg Ser
                800                 805                 810
Met Gly Ala Lys Val Val Ile Ala Ile Asp Val Gly Ser Arg Asp
                815                 820                 825
Glu Thr Asp Leu Thr Asn Tyr Gly Asp Ala Leu Ser Gly Trp Trp
                830                 835                 840
Leu Leu Trp Lys Arg Trp Asn Pro Leu Ala Thr Lys Val Lys Val
```

```
                845                 850                 855
Leu Asn Met Ala Glu Ile Gln Thr Arg Leu Ala Tyr Val Cys Cys
                860                 865                 870
Val Arg Gln Leu Glu Val Val Lys Ser Ser Asp Tyr Cys Glu Tyr
                875                 880                 885
Leu Arg Pro Pro Ile Asp Ser Tyr Ser Thr Leu Asp Phe Gly Lys
                890                 895                 900
Phe Asn Glu Ile Cys Glu Val Gly Tyr Gln His Gly Arg Thr Val
                905                 910                 915
Phe Asp Ile Trp Gly Arg Ser Gly Val Leu Glu Lys Met Leu Arg
                920                 925                 930
Asp Gln Gln Gly Pro Ser Lys Lys Pro Ala Ser Ala Val Leu Thr
                935                 940                 945
Cys Pro Asn Ala Ser Phe Thr Asp Leu Ala Glu Ile Val Ser Arg
                950                 955                 960
Ile Glu Pro Ala Lys Pro Ala Met Val Asp Asp Glu Ser Asp Tyr
                965                 970                 975
Gln Thr Glu Tyr Glu Glu Glu Leu Leu Asp Val Pro Arg Asp Ala
                980                 985                 990
Tyr Ala Asp Phe Gln Ser Thr Ser Ala Gln Gln Gly Ser Asp Leu
                995                1000                1005
Glu Asp Glu Ser Ser Leu Arg His Arg His Pro Ser Leu Ala Phe
               1010                1015                1020
Pro Lys Leu Ser Glu Gly Ser Ser Asp Gln Asp Gly
               1025                1030

<210> 10
<211> 407
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7491216CD1

<400> 10
Met Trp Asp Leu Ala Leu Ile Phe Leu Ala Ala Ala Cys Val Phe
  1               5                  10                  15
Ser Leu Gly Val Thr Leu Trp Val Ile Cys Ser His Phe Phe Thr
                20                  25                  30
Val His Ile Pro Ala Ala Val Gly His Pro Val Lys Leu Arg Val
                35                  40                  45
Leu His Cys Ile Phe Gln Leu Leu Leu Thr Trp Gly Met Ile Phe
                50                  55                  60
Glu Lys Leu Arg Ile Cys Ser Met Pro Gln Phe Phe Cys Phe Met
                65                  70                  75
Gln Asp Leu Pro Pro Leu Lys Tyr Asp Pro Asp Val Val Val Thr
                80                  85                  90
Asp Phe Arg Phe Gly Thr Ile Pro Val Lys Leu Tyr Gln Pro Lys
                95                 100                 105
Ala Ser Thr Cys Thr Leu Lys Pro Gly Ile Val Tyr Tyr His Gly
               110                 115                 120
Gly Gly Gly Val Met Gly Ser Leu Lys Thr His His Gly Ile Cys
               125                 130                 135
Ser Arg Leu Cys Lys Glu Ser Asp Ser Val Val Leu Ala Val Gly
               140                 145                 150
Tyr Arg Lys Leu Pro Lys His Lys Phe Pro Val Pro Val Arg Asp
               155                 160                 165
Cys Leu Val Ala Thr Ile His Phe Leu Lys Ser Leu Asp Ala Tyr
               170                 175                 180
Gly Val Asp Pro Ala Arg Val Val Val Cys Gly Asp Ser Phe Gly
               185                 190                 195
Gly Ala Ile Ala Ala Val Val Cys Gln Gln Leu Val Asp Arg Pro
```

```
                200                 205                 210
Asp Leu Pro Arg Ile Arg Ala Gln Ile Leu Ile Tyr Ala Ile Leu
                215                 220                 225
Gln Ala Leu Asp Leu Gln Thr Pro Ser Phe Gln Gln Arg Lys Asn
                230                 235                 240
Ile Pro Leu Leu Thr Trp Ser Phe Ile Cys Tyr Cys Phe Phe Gln
                245                 250                 255
Asn Leu Asp Phe Ser Ser Ser Trp Gln Glu Val Ile Met Lys Gly
                260                 265                 270
Ala His Leu Pro Ala Glu Val Trp Glu Lys Tyr Arg Lys Trp Leu
                275                 280                 285
Gly Pro Glu Asn Ile Pro Glu Arg Phe Lys Glu Arg Gly Tyr Gln
                290                 295                 300
Leu Lys Pro His Glu Pro Met Asn Glu Ala Ala Tyr Leu Glu Val
                305                 310                 315
Ser Val Val Leu Asp Val Met Cys Ser Pro Leu Ile Ala Glu Asp
                320                 325                 330
Asp Ile Val Ser Gln Leu Pro Glu Thr Cys Ile Val Ser Cys Glu
                335                 340                 345
Tyr Asp Ala Leu Arg Asp Asn Ser Leu Leu Tyr Lys Lys Arg Leu
                350                 355                 360
Glu Asp Leu Gly Val Pro Val Thr Trp His His Met Glu Asp Gly
                365                 370                 375
Phe His Gly Val Leu Arg Thr Ile Asp Met Ser Phe Leu His Phe
                380                 385                 390
Pro Cys Ser Met Arg Ile Leu Ser Ala Leu Val Gln Phe Val Lys
                395                 400                 405
Gly Leu
```

<210> 11
<211> 352
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71624817CD1

<400> 11

```
Met Ser Leu Tyr Cys Gly Ile Ala Cys Arg Arg Lys Phe Phe Trp
  1               5                  10                  15
Cys Tyr Arg Leu Leu Ser Thr Tyr Val Thr Lys Thr Arg Tyr Leu
                 20                  25                  30
Phe Glu Leu Lys Glu Asp Asp Asp Ala Cys Lys Lys Ala Gln Gln
                 35                  40                  45
Thr Gly Ala Phe Tyr Leu Phe His Ser Leu Ala Pro Leu Leu Gln
                 50                  55                  60
Thr Ser Ala His Gln Tyr Leu Ala Pro Arg His Ser Leu Leu Glu
                 65                  70                  75
Leu Glu Arg Leu Leu Gly Lys Phe Gly Gln Asp Ala Gln Arg Ile
                 80                  85                  90
Glu Asp Ser Val Leu Ile Gly Cys Ser Glu Gln Gln Glu Ala Trp
                 95                 100                 105
Phe Ala Leu Asp Leu Gly Leu Asp Ser Ser Phe Ser Ile Ser Ala
                110                 115                 120
Ser Leu His Lys Pro Glu Met Glu Thr Glu Leu Lys Gly Ser Phe
                125                 130                 135
Ile Glu Leu Arg Lys Ala Leu Phe Gln Leu Asn Ala Arg Asp Ala
                140                 145                 150
Ser Leu Leu Ser Thr Ala Gln Ala Leu Leu Arg Trp His Asp Ala
                155                 160                 165
His Gln Phe Cys Ser Arg Ser Gly Gln Pro Thr Lys Lys Asn Val
```

```
                170                 175                 180
Ala Gly Ser Lys Arg Val Cys Pro Ser Asn Asn Ile Ile Tyr Tyr
                185                 190                 195
Pro Gln Met Ala Pro Val Ala Ile Thr Leu Val Ser Asp Gly Thr
                200                 205                 210
Arg Cys Leu Leu Ala Arg Gln Ser Ser Phe Pro Lys Gly Met Tyr
                215                 220                 225
Ser Ala Leu Ala Gly Phe Cys Asp Ile Gly Glu Ser Val Glu Glu
                230                 235                 240
Thr Ile Arg Arg Glu Val Ala Glu Glu Val Gly Leu Glu Val Glu
                245                 250                 255
Ser Leu Gln Tyr Tyr Ala Ser Gln His Trp Pro Phe Pro Ser Gly
                260                 265                 270
Ser Leu Met Ile Ala Cys His Ala Thr Val Lys Pro Gly Gln Thr
                275                 280                 285
Glu Ile Gln Val Asn Leu Arg Glu Leu Glu Thr Ala Ala Trp Phe
                290                 295                 300
Ser His Asp Glu Val Ala Thr Ala Leu Lys Arg Lys Gly Pro Tyr
                305                 310                 315
Thr Gln Gln Gln Asn Gly Thr Phe Pro Phe Trp Leu Pro Pro Lys
                320                 325                 330
Leu Ala Ile Ser His Gln Leu Ile Lys Glu Trp Val Glu Lys Gln
                335                 340                 345
Thr Cys Ser Ser Leu Pro Ala
                350
```

<210> 12
<211> 385
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6945964CD1

<400> 12

```
Met Glu Glu Glu Lys Asp Asp Ser Pro Gln Ala Asp Phe Cys Leu
  1                 5                  10                  15
Gly Thr Ala Leu His Ser Trp Gly Leu Trp Phe Thr Glu Glu Gly
                 20                  25                  30
Ser Pro Ser Thr Met Leu Thr Gly Ile Ala Val Gly Ala Leu Leu
                 35                  40                  45
Ala Leu Ala Leu Val Gly Val Leu Ile Leu Phe Met Phe Arg Arg
                 50                  55                  60
Leu Arg Gln Phe Arg Gln Ala Gln Pro Thr Pro Gln Tyr Arg Phe
                 65                  70                  75
Arg Lys Arg Asp Lys Val Met Phe Tyr Gly Arg Lys Ile Met Arg
                 80                  85                  90
Lys Val Thr Thr Leu Pro Asn Thr Leu Val Glu Asn Thr Ala Leu
                 95                 100                 105
Pro Arg Gln Arg Ala Arg Lys Arg Thr Lys Val Leu Ser Leu Ala
                110                 115                 120
Lys Arg Ile Leu Arg Phe Lys Lys Glu Tyr Pro Ala Leu Gln Pro
                125                 130                 135
Lys Glu Pro Pro Pro Ser Leu Leu Glu Ala Asp Leu Thr Glu Phe
                140                 145                 150
Asp Val Lys Asn Ser His Leu Pro Ser Glu Val Leu Tyr Met Leu
                155                 160                 165
Lys Asn Val Arg Val Leu Gly His Phe Glu Lys Pro Leu Phe Leu
                170                 175                 180
Glu Leu Cys Lys His Ile Val Phe Val Gln Leu Gln Glu Gly Glu
                185                 190                 195
His Val Phe Gln Pro Arg Glu Pro Asp Pro Ser Ile Cys Val Val
```

```
                200                 205                 210
Gln Asp Gly Arg Leu Glu Val Cys Ile Gln Asp Thr Asp Gly Thr
                215                 220                 225
Glu Val Val Val Lys Glu Val Leu Ala Gly Asp Ser Val His Ser
                230                 235                 240
Leu Leu Ser Ile Leu Asp Ile Ile Thr Gly His Ala Ala Pro Tyr
                245                 250                 255
Lys Thr Val Ser Val Arg Ala Ala Ile Pro Ser Thr Ile Leu Arg
                260                 265                 270
Leu Pro Ala Ala Ala Phe His Gly Val Phe Glu Lys Tyr Pro Glu
                275                 280                 285
Thr Leu Val Arg Val Val Gln Ile Ile Met Val Arg Leu Gln Arg
                290                 295                 300
Val Thr Phe Leu Ala Leu His Asn Tyr Leu Gly Leu Thr Thr Glu
                305                 310                 315
Leu Phe Asn Ala Glu Ser Gln Ala Ile Pro Leu Val Ser Val Ala
                320                 325                 330
Ser Val Ala Ala Gly Lys Ala Lys Lys Gln Val Phe Tyr Gly Glu
                335                 340                 345
Glu Glu Arg Leu Lys Lys Pro Pro Arg Leu Gln Glu Ser Cys Asp
                350                 355                 360
Ser Gly Thr Val Leu His Gln Gly Gly Gln Cys Pro Ala Pro Glu
                365                 370                 375
Trp Gly Met Glu Trp Thr Glu Trp Ser Glu
                380                 385

<210> 13
<211> 1437
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2425607CB1

<400> 13
tgcgacggcg ggctggcctg gcgcgctgcg gacgtcgctc acccgctccc gaggaagggc 60
agtgggcccc gccgccgcct cccaatggcg aggctgcggg attgcctgcc ccgcctgatg 120
ctcacgctcc ggtccctgct cttctggtcc ctggtctact gctactgcgg gctctgcgcc 180
tccatccacc tgctcaaact tttgtggagc ctcggcaagg ggccgcgcag accttccggc 240
ggcccgcccg ggagcaccct ccccgcgtgc ctgagcgacc cctccttggg cacccactgc 300
tacgtgcgga tcaaggattc agggttaaga tttcactatg ttgctgctgg agaaagaggc 360
aaaccactta tgctgctgct tcatggattt cccgaattct ggtatatctc tggcggtaac 420
caactgagag aatttaaaag tgaatatcga gttgtagcac tggatttgag aggttatgga 480
gaaacagatg ctcccattca tcgacagaat tataaattgg attgtctaat tacagatata 540
aaggatattt tagattcttt agggtatagc aaatgtgttc ttattggcca tgactggggg 600
ggcatgattg cttggctaat tgccatctgt tatcctgaaa tggtgatgaa gcttattgtt 660
attaacttcc ctcatccaaa tgtatttaca gaatatattt tacgacaccc tgctcagctg 720
ttgaaatcca gttattatta cttcttccaa ataccatggt tcccagaatt tatgttctca 780
ataaatgatt tcaaggtttt gaaacatctg tttaccagtc acagcactgg cattggaaga 840
aaaggatgcc aattaacaac agaggatctt gaagcttata tttatgtctt ttctcagcct 900
ggagcattaa gtggcccaat taaccattac cgaaatatct tcagctgcct gcctctcaaa 960
catcacatgg tgaccactcc aacactacta ctgtggggag agaatgacgc attcatggag 1020
gttgagatgg ctgaagtcac aaagatttat gttaaaaact atttcaggct aactattttg 1080
tcagaagcca gtcattggct tcagcaagac caacctgaca tagtgaacaa attgatatgg 1140
acatttctaa aagaagaaac aagaaaaaaa gattgacttt tctttatctt ctatgaaggg 1200
tctgtaatga aatctctaaa taatttttaa aaattgttca tcaacttctt tatgttttat 1260
tagaaaaaaa ctgttttaat gtgctttatc ataaataaat atcctgacaa atggtattga 1320
aaaaaatctg agatcatgtg aacatataat acaatgttga ttttcttctg gtgtattttg 1380
cacagacaag catctgcctt aaaatatata cacttgtaca aaaaggaaaa aaaaaaa    1437

<210> 14
<211> 3269
```

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2786919CB1

<400> 14
cgccgccgcc gccgctggca aaaagcggag cccagggaag cgtgtcctgc tcagaccgcc 60
ttccttctcc attgccaccc gtgccgaaca gccaggctgc ccaattgcaa ctgtagacca 120
atgaactaat ccatcgcccg cagcccgact ctcagcagcg gttcgtcccg gtgcccaccc 180
ccgcgaagcg gagcgcccgg gcacctagcc tccctgccgg ccacctagcc tccctgccgg 240
ccacgatgcc gaacgtgctg ctgccgccca aggagagcaa cctcttcaaa cgcatcttga 300
aatgttatga acagaagcag tacaaaaatg gcctcaagtt ttgcaagatg attctgtcga 360
acccaaaatt tgctgaacat ggagagactt tggctatgaa aggattaaca ctgaactgtt 420
taggaaaaaa agaagaagct tatgagtttg ttcgtaaagg acttcgtaat gatgtcaaga 480
gtcatgtctg ttggcatgta tatggactct tgcagcgttc tgataaaaaa tatgatgaag 540
ctataaaatg ttaccgaaat gccctcaaat tagataaaga taacctgcaa attttgaggg 600
atctctcact gttgcagatc caaatgagag accttgaagg ttaccgagag acaagatacc 660
agcttcttca gttgcgcccc acacagcgtg cctcctggat tggatatgct attgcatacc 720
atttgctgaa agattatgat atggccctaa aactgttgga agaatttaga caaactcagc 780
aagttcctcc aaacaaaata gattatgaat atagtgaatt gatattatac cagaatcaag 840
tgatgagaga ggcagatctg ttgcaggaat ctttggaaca tatagaaatg tatgagaaac 900
aaatatgtga taaacttttg gtggaagaaa ttaaagggga aatactgttg aaattgggaa 960
gattaaaaga agccagtgaa gtgttcaaaa acttgattga tcgaaatgca gaaaactggt 1020
gttattatga aggcttggaa aaagctctac aaattagcac tttagaagag aggcttcaaa 1080
tttatgaaga aattagtaag cagcacccca aagcaattac acccagaaga ttacctttga 1140
ctcttgtccc aggtgaaaga tttagagaac taatggataa gttcctgagg gttaacttca 1200
gtaaaggctg cccacccttg tttactactt tgaaatcttt atattacaat acagaaaagg 1260
tttctataat ccaggaactt gttactaatt atgaagcctc tcttaaaacg tgtgactttt 1320
ttagcccata tgagaatggg gagaaggaac ccccgacaac actactctgg gttcagtatt 1380
tcctggcaca gcactttgat aaacttggac agtattcttt ggctttggat tatattaatg 1440
ctgcaattgc tagtactcca actctaatag aattattcta tatgaaagca aaaatttaca 1500
agcatatagg taatctcaaa gaagctgcaa agtggatgga tgaagcacag tctttggaca 1560
cagctgatag attcatcaat tccaaatgtg caaaatacat gcttcgagca aatatgataa 1620
aagaagcaga ggaaatgtgc tccaagttca caagggaagg aacatctgcc atggaaaatc 1680
taaatgaaat gcagtgtatg tggtttcaga cagaatgcat ttcagcttat cagcgtctgg 1740
ggagatacgg ggatgccttg aaaaaatgtc atgaagtaga aaggcatttt tttgagataa 1800
ctgatgacca attcgacttc catacatact gcatgagaaa gatgaccctt cgtgcctatg 1860
ttgacctttt gagattagaa gatatactca gaagacatgc cttttatttc aaggctgcta 1920
gatcagcgat tgaaatatac ttgaaattgt atgataatcc cttaaccaat gaaagcaaac 1980
aacaagaaat aaactcagaa aacttgtcag ccaaagaatt gaagaaaatg cttagcaagc 2040
agagaagagc tcagaaaaag gctaaactag aagaagaaag aaagcatgca gaaagagaac 2100
gtcaacagaa aaatcaaaag aaaaaaagag atgaagaaga agaagaagcc agtggcctta 2160
aggaagaact tatacctgaa aaattagaaa gggtagaaaa tccattagag gaagccgtta 2220
agttccttat acctcttaag aaccttgttg ctgataacat tgacactcat ctgttagcat 2280
ttgaaatata ttttagaaaa ggaaagtttc tgttaatgct gcagtctgtc aaacgagctt 2340
ttgccattaa cagtaataac ccatggttac atgaatgttt aattagattt tctaaatctg 2400
tgtctaatca tagtaatctt ccagacattg tgagcaaagt tctatctcaa gaaatgcaga 2460
aaatatttgt caaaaaggat ttggaaagtt ttaatgagga ttttctgaaa cgtaacgcta 2520
cctctcttca gcatctactt tcaggtgcta aaatgatgta ttttctggac aagtcaaggc 2580
aggagaaagc aattgctata gccactagac tagatgaaac tataaaagat aaagatgtaa 2640
agacattaat aaaggtttct gaagcactgc ttgatggcag ctttgggaac tgtagttccc 2700
aatatgaaga atataggatg gcctgtcata acctgcttcc ttttacatct gccttcttgc 2760
ctgctgtgaa tgaagtcgac aatcctaatg tggcactgaa ccatacagct aattatgatg 2820
tcttggcaaa tgaaatttga aatcttctag aaagtagact cctatagact caaagctgaa 2880
ttgagatcag ggtttctttt ccaggttgca ttttaatata cgtatgaaat gaaatatttg 2940
gttaggattt ttaaatggca tattctgtaa gcttattttg ttctttaccc gacctgccaa 3000
tttacataac catctgttaa aattacctgt ttattcttac acagttttgt ggtagctccg 3060
atcgcttctg tataattata atttcttact aagctaagtt gtatacaacc ccactgtatt 3120
attttcttta gattctgatt tcaggatctg tgctttgatt ccttgtctgt tttatcgtat 3180
atttgcttgc tttttaatta aagcacatca gttttaaagt gttaactata tgttaacaat 3240
atggaggctg ggcgtggtgg ctcacacct 3269

<210> 15
<211> 1534
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1801130CB1

<220>
<221> unsure
<222> 91
<223> a, t, c, g, or other

<400> 15
atggaggaaa aaggaataaa gaggatttca aaaccgctag gagtgaagcg agaaggtaaa 60
ggatataaag aatataagag gtttcaggct ncaggtcctg gcaggaggat cgcgggccag 120
aggcggagcc accatcctac cgcgactttc agactccgac cgtggcctcg cgttggtgac 180
gggggcggct cggctgttgc tggagaccgg cggcgtgaga ccgctccccg agccttgcgg 240
ggccgccggg tccgcgcgcc gctgccgatg tccgcgcagg ttcataggca gaagggactt 300
gacttgtctc agataccta ttttaatctt gtgaagcatt taacccctgc ctgtccaaat 360
gtatatagta tatcacagtt tcatcacaca accccagaca gtaaaacaca cagtggtgaa 420
aaatacaccg atcctttcaa actcggttgg agagacttga aaggtctgta tgaggacatt 480
agaaaggaac tgcttatatc aacatcagaa cttaaggaaa tgtctgagta ctactttgat 540
gggaaaggga aagcctttcg accaattatt gtggcgctaa tggcccgagc atgcaatatt 600
catcataaca actcccgaca tgtgcaagcc agccagcgcg ccatagcctt aattgcagaa 660
atgatccaca ctgctagtct ggttcacgat gacgttattg acgatgcaag ttctcgaaga 720
ggaaaacaca cagttaataa gatctggggt gaaaagaagg ctgttcttgc tggagattta 780
attctttctg cagcatctat agctctggca cgaattggaa atacaactgt tatatctatt 840
ttaacccaag ttattgaaga tttggtgcgt ggtgaatttc ttcagctcgg gtcaaaagaa 900
aatgagaatg aaagatttgc acactacctt gagaagacat tcaagaagac cgccagcctg 960
atagccaaca gttgtaaagc agtctctgtt ctaggatgtc ccgacccagt ggtgcatgag 1020
atcgcctatc agtacggaaa aaatgtagga atagctttc agctaataga tgatgtattg 1080
gacttcacct cgtgttctga ccagatgggc aaaccaacat cagctgatct gaagctcggg 1140
ttagccactg gtcctgtcct gtttgcctgt cagcagttcc cagaaatgaa tgctatgatc 1200
atgcgacggt tcagtttgcc tggagatgta gacagagctc gacagtatgt actacagagt 1260
gatggtgtgc aacaaacaac ctacctcgcc cagcagtact gccatgaagc aataagagag 1320
atcagtaaac ttcgaccatc cccagaaaga gatgccctca ttcagctttc agaaattgta 1380
ctcacaagag ataaatgaca actctttctg ttctttctgg cagctattct taccagactg 1440
tgcctaaaga attttgtgga atacactttg tttgcttcat gtgcagatga ccaaaaatca 1500
ttttaaagat atcaaactta ttgatgggca ttta 1534

<210> 16
<211> 3400
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3535146CB1

<400> 16
ggcacgcgcc ccgccgccct cggcagccgc agctccgtgt ccctgagaac cagccgtccc 60
gcgccatggg cacgcgtctg ccgctcgtcc tgcgccagct ccgccgcccg ccccagcccc 120
cgggccctcc gcgccgcctc cgtgtgccct gtcgcgctag cagcggcggc ggcggaggcg 180
gcggcggtgg ccgggagggc ctgcttggac agcggcggcc gcaggatggc caggcccgga 240
gcagctgcag ccccggcggc cgaacgcccg cggcgcggga ctccatcgtc agagaagtca 300
ttcagaattc aaaagaagtt ctaagtttat tgcaagaaaa aaaccctgcc ttcaagccgg 360
ttcttgcaat tatccaggca ggtgacgaca acttgatgca ggaaatcaac cagaatttgg 420
ctgaggaggc tggtctgaac atcactcaca tttgcctccc tccagatagc agtgaagccg 480
agattataga tgaaatctta aagatcaatg aagataccag agtacatggc cttgcccttc 540
agatctctga gaacttgttt agcaacaaag tcctcaatgc cttgaaacca gaaaaagatg 600
tggatggagt aacagacata aacctgggga agctggtgcg aggggatgcc catgaatgtt 660

ttgtttcacc tgttgccaaa gctgtaattg aacttcttga aaaatcaggt gtcaacctag 720
atggaaagaa gattttggta gtgggggccc atgggtcttt ggaagctgct ctacaatgcc 780
tgttccagag aaaagggtcc atgacaatga gcatccagtg gaaaacacgc cagcttcaaa 840
gcaagcttca cgaggctgac attgtggtcc taggctcacc taagccagaa gagattcccc 900
ttacttggat acaaccagga actactgttc tcaactgctc ccatgacttc ctgtcaggga 960
aggttgggtg tggctctcca agaatacatt ttggtggact cattgaggaa gatgatgtga 1020
ttctccttgc tgcagctctg cgaattcaga acatggtcag tagtggaagg agatggcttc 1080
gtgaacagca gcacaggcgg tggagacttc actgcttgaa acttcagcct ctctcccctg 1140
tgccaagtga cattgagatt tcaagaggac aaactccaaa agctgtggat gtccttgcca 1200
aggagattgg attgcttgca gatgaaattg aaatctatgg caaaagcaaa gccaaagtac 1260
gtttgtccgt gctagaaagg ttaaaggatc aagcagatgg aaaatacgtc ttagttgctg 1320
ggatcacacc cacccctctt ggagaaggga agagcacagt caccatcggg cttgtgcagg 1380
ctctgaccgc acacctgaat gtcaactcct ttgcctgctt gaggcagcct tcccaaggac 1440
cgacgtttgg agtgaaagga ggagccgcgg gtggtggata tgcccaggtc atccccatgg 1500
aggagttcaa ccttcacttg actggagaca tccacgccat caccgctgcc aataacttgc 1560
tggctgccgc catcgacacg aggattcttc atgaaaacac gcaaacagat aaggctctgt 1620
ataatcggct ggttccttta gtgaatggtg tcagagaatt ttcagaaatt cagcttgctc 1680
ggctaaaaaa actgggaata aataagactg atccgagcac actgacagaa gaggaagtga 1740
gtaaatttgc ccgtctcgac atcgacccat ctaccatcac gtggcagaga gtattggata 1800
caaatgaccg atttctacga aaaataacca tcgggcaggg aaacacagag aagggccatt 1860
accggcaggc gcagtttgac atcgcagtgg ccagcgagat catggcggtg ctggccctga 1920
cggacagcct cgcagacatg aaggcacggc tgggaaggat ggtggtggcc agtgacaaaa 1980
gcgggcagcc tgtgacagca gatgatttgg gggtgacagg tgctttgaca gttttgatga 2040
aagatgcaat aaaaccaaac ctgatgcaga ccctggaagg gacacctgtg ttcgtgcatg 2100
cgggcccttt tgctaacatt gctcacggca actcttcagt gttggctgat aaaattgccc 2160
tgaaactggt tggtgaagaa ggatttgtag tgaccgaagc tggctttggt gctgacatcg 2220
gaatggagaa attcttcaac atcaagtgcc gagcttccgg cttggtgccc aacgtggttg 2280
tgttagtggc aacggtgcga gctctgaaga tgcatggagg cgggccaagt gtaacggctg 2340
gtgttcctct taagaaagaa tatacagagg agaacatcca gctggtggca gacggctgct 2400
gtaacctcca gaagcaaatt cagatcactc agctctttgg ggttcccgtt gtggtggctc 2460
tgaatgtctt caagaccgac acccgcgctg agattgactt ggtgtgtgag cttgcaaagc 2520
gggctggtgc ctttgatgca gtcccctgct atcactggtc ggttggtgga aaaggatcgg 2580
tggacttggc tcgggctgtg agagaggctg cgagtaaaag aagccgattc cagttcctgt 2640
atgatgttca ggttccaatt gtggacaaga taaggaccat tgctcaggct gtctatggag 2700
ccaaagatat tgaactctct cctgaggcac aagccaaaat agatcgttac actcaacagg 2760
gttttggaaa tttgcccatc tgcatggcaa agacccacct ttctctatct caccaacctg 2820
acaaaaaagg tgtgccaagg gacttcatct tacctatcag tgacgtccgg gccagcatag 2880
gcgctgggtt catttaccct ttggtcggaa cgatgagcac catgccagga ctgcccaccc 2940
ggccctgctt ttatgacata gatcttgata ccgaaacaga acaagttaaa ggcttgttct 3000
aagtggacaa ggctctcaca ggacccgatg cagactcctg aaacagacta ctctttgcct 3060
ttttgctgca gttggagaag aaactgaatt tgaaaaatgt ctgttatgca atgctggaga 3120
catggtgaaa taggccaaag atttcttctt cgttcaagat gaattctgtt cacagtggag 3180
tatggtgttc ggcaaaagga cctccaccaa gactgaaaga aactaattta tttctgtttc 3240
tgtggagttt ccattatttc tactgcttac actttagaat gtttatttta tggggactaa 3300
gggattaaga gtgtgaacta aaaggtaaca ttttccactc tcaagttttc tactttgtct 3360
ttgaactgaa aataaacatg gatctagaaa accaaaaaaa 3400

<210> 17
<211> 3395
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1436543CB1

<400> 17
caggtctggg ctacaaaagt atggccgctt ctgaggcggc ggtggtgtct tcgccgtctt 60
tgaaaacaga cacatcccct gtccttgaaa ctgcaggaac ggtcgcagca atggctgcga 120
ccccgtcagc aagggctgca gccgcggtgg ttgcggccgc ggccaggacc ggatccgaag 180
ccagggtctc caaggccgct ttggctacca agctgctgtc cttgagcggc gtgttcgccg 240
tgcacaagcc caaagggccc acttcagccg agctgctgaa tcggttgaag gagaagctgc 300
tggcagaagc tggaatgcct tctccagaat ggaccaagag gaaaaagcag actttgaaaa 360

ttgggcatgg agggactcta gacagcgcag cccgaggagt tctggttgtt ggaattggaa 420
gcggaacaaa aatgttgacc agtatgttgt cagggtccaa gagatatact gccattggag 480
aactggggaa agctactgat acactagatt ctacggggag ggtaacagaa gaaaaacctt 540
acgataaaat aacacaagaa gatattgaag gcattctaca gaaatttact ggaaatataa 600
tgcaagtgcc ccccctctat tctgcattaa agaaagatgg acaaagactt tcgactttga 660
tgaagagagg tgaagtcgta gaagcaaaac ctgccaggcc agtgactgta tacagtatct 720
cccttcaaaa attccagcca ccatttttca cattagatgt tgaatgtgga ggaggttttt 780
atatcagaag cttggtcagt gacattggaa aagaactatc ttcctgtgcc aatgtgctag 840
agctgacccg aaccaaacag ggaccattta cgctagaaga acatgccctt cctgaagaca 900
aatggacaat tgatgacatt gcacagtctc ttgagcattg ctcatctctt ttcccagcag 960
agttggcact taaaaaatca aaacctgagt ctaatgaaca ggttttgagc tgtgaatata 1020
taactctaaa tgagccaaag agagaagatg atgtaattaa gacgtgttga gattggcctg 1080
ggaatatcat cattttctag ttgacatttg aatcctgtgt gcagatgcag aatgacaagc 1140
tgcattcaaa agacaaacaa tatgtctttt ttttttttgc atgaagaaaa atgtctatca 1200
tttacagttt caatagcaca taatttattt tctatgcatt ataaatggcc ttgcagttgg 1260
ctcagttgtt tgctgtgttg tgaaatgttt taggattttt tgtattgtga aaatatgaat 1320
atgattggat tcagaaaaat taacttttctg aatttgatct gtcttcagtc ttgtgaaaaa 1380
gttgaacaaa tttcctaatc aaagaaaaaa gtatgagctc catgtttctt tagtttcaca 1440
aaaatgacca taatttagtg ttatttttac tttatttagg cttcctggtg gcttcatttt 1500
attgaaattc tttaaattgt ttaaagtggc cattattgat ctctttcttc tgttttggag 1560
agtttattat taaaaacatt tctttgataa aatggccatc atctagtaat acctgtgttt 1620
gtttagatct tggaaatgaa taagctttga taatatttgt aaatgaacca aattattact 1680
gctaccacta acaggttgta aatagaagac taatacttaa ttaaagtcac cttcctacca 1740
ttagagcaga agacagctcc tatagttttg tattttggca gctatgagat attttcatgg 1800
taatgtcaac atggtcaagc actttgtacc aagttattaa gtaacataat ttttaaaatt 1860
taaagaatgt gtcttcaact aaaaacttta ttctttagca tttatttata tttctctgta 1920
gggtgttccc tgtgacattg tctctttagt ttgctctttc aagagatact tacagatgtt 1980
gagatggctg ccctgcattt ccagctaatc tcttctgctc taaatattta aaaacagttc 2040
ttctcaaaca ttttcattca gatagctttc tgaaagttcc ctatccctct ttaccataat 2100
tttttaaatg tagccacatt gtaatagtaa acttcatata taatgagtgc ttcatatttt 2160
tgttatggga aagcaatata ttatgcagcc agtctgtaga aacattcaga tccctcttcc 2220
tttactcaaa tacagtttca aaaggaagac tcatgagaaa tttcataaaa tacaagtttt 2280
tagatgttta tgctttgcct ttctttttaa aggtgttttc ctgctttgta gtctctaact 2340
ctgaaattta aaatatgtaa actaaagtgg ttttatttgt gcttaaccca atttaaactc 2400
aatgtaaaat gttatatatg catcagtaca gcattttcaa catattggca acatatttta 2460
aatgaaaaca ctaaaacaat tcttagtatg agacaaaact gtaaggaaaa agagtgttaa 2520
taccatgatg cattaacata aaatatcaaa cacacaaagt cataaaatga aaatttacag 2580
ttttacctgt tcatatctag tgccccacag tgtgtgtcaa ccaaaggtgg cagtggctac 2640
atctgcctgt tggactggta caggttacaa tatgtcctct tccattgcaa attaaagtcc 2700
aaatagagaa atacttaggt tttagaacac atcagaggta tttctgctgt atttttcacc 2760
ttaaaaattg acacagagtt tactaataga ggagtagaga ttgttgacca tttttaaaaa 2820
acgatagcca ctctttttct tttatgttta aaactgaagt tttgccaaat gggaaaatta 2880
ctgttacctc taccatctta atgtagtaac tttagaattt aaatttttat attactattt 2940
tcctttttgt tgttcacata gtcttaaggc acctatactt ttaaattgac tttttcattt 3000
gatattatct atatgtatgt agttgtgata atgattattt taattatatt actttatact 3060
cttaatttat ttagagtatt tctctattgc tgaatactta agtagtttta aattttatta 3120
tgataaattc ctgggagggg gattatttag tgaaataata tgaagaactt tatgacttat 3180
gtttgcctta ttgcattccc aaagagttgt aacattttac agtgttacca tttgagtagg 3240
ggttttatat gttgttgcta atttagtaaa cataggagag aaatcaaagt ttttctgatt 3300
tgcttttatg tgatttatct gtatactttg ttcatttata taaataaatg tcttaatggt 3360
ttctatacat aaaaaaaaaa aaaaaaaaaa aaaaa 3395

<210> 18
<211> 2283
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7491063CB1

<400> 18
gtactatttt gccgcttcct gtgagtctat aattatttca aaataaaaag attatttaaa 60

aaagaatgtg ggggccagga gcattggctc atgcctataa tccaagcact ttgggaggct 120
gaggtgggcg gattgcttga gctcaggagt ttgagaccag cctgggcaac atggcaaaac 180
cctgtctcta aaaaaagaaa aaagaaaaag aatgtgggga aaaatcattt ggagacttaa 240
aatgtagaaa caaagaacat aaatgtggta taagaagttt taggtgacaa tatagttttg 300
aggctaaacc aaaaaggaaa aggagatata tctgactgta tatctaagaa ctcaccaata 360
agaaagacgt gggccctgct ctgttgcctg cctttgctcc ccaggccccg agtggttcac 420
tgagccatga agacagattc caggcaccag gaattgcacc tccaatccca gatgctatgt 480
tcagcaaagg ctccatggtt ctggcctaca gtgccggcct ggacacctcc tgcatcctcg 540
tgtggctgaa ggaacaaggc tatgacatta ttgcctacct ggccaatgtt ggccagaagg 600
aagacttcaa ggaagccagg aagaaggcac tgaaccttgg gtccaaaaag gtgttcattg 660
aggacgtcag caaggagttc gtggaggagt tcatctggcc ttatgaggac tgctacctcc 720
tgggcccctc tcttcccagg ccttacatca cccgcaaaca agtggaaatt gctcagtggg 780
agggggcaaa gtatgtgtcc cacagtgcca tgggaaaggg gaacgatcag gtctggtttg 840
agctcgcctg ctactcgctg gcccccaga ttaaggtcat tgctcccggg aggatccctg 900
agttctacaa ccagtccaag ggccgcagtg atctgatgga atatgcagag aaacatggga 960
ttcccatccc agtcactctg aagcacccat ggaacatgga cgagaacctc atgcacatca 1020
gccacgaggg ctggaatctt ggagaacccc aagaaccaga ggcaccttca ggtctctaca 1080
tgaagattca ggacctggcc aaagccccca acacccccaa cattttcaag actgaggggg 1140
tccctgtgaa ggtgaccagc atcaaggatg gcaccaccca ccagacctcc ttggtgctct 1200
tacatgtaac ctggaatgaa gtcgcaggca agcacagcgt gggccatatt gacatcgtgg 1260
agaaccgctt cattgaaatg aatatctgca agaccccagc aggcaccatc ctttaccacc 1320
ctcatttaga cattgagggc tttgccatgg aacaggaagt gcgcaaaatc aaacaaggcc 1380
tgggcttgaa atttgctgag ttggtgtaca ccggtttctg gcacaaccct cagtgtgact 1440
ttgcccacca ctgcattgcc aagtcccagg accgagtgga agggaaagtg caggtatcca 1500
tcttcaaggg ccaggtgtac atcctctgcc aggagccccc actgtctctc tacagtggag 1560
agcaggtgag catgaacgtg gagggcaatg atgagccagc cagtcgtctc atcaacatca 1620
attccctcag gatgaaggaa tatcatcatc tccagagcaa ggtcactgcc aaatagaccc 1680
ctgtacaatg gggagctagg gccacctcac tttgcagatt ccccaagtac aggcactaat 1740
tgttgtgata atttgtaatt gtgacttgtt ctccctggct aagagtgtag tggggctgcc 1800
gggccccagc tttgttccct ggtccccctg aaatggtcat caaagggaag ggtgaggggc 1860
agctacagtg gggagctata aaatgacaat taaaaaaaaa aaaaaaaaag actaaggcca 1920
gccacggggg ctcacgtcgg taatcccggc acttggggag accgaggtgg gcaagatcac 1980
cggaggtcag gcagtctggc gcaccagccg ggccacaacg tggggaaacc ccatctctac 2040
gtaaagcaat cccacaaatt agcggggcgc ggtgttgtgt gcctggtaaa cccagcaaat 2100
ggcgagggct tgaggccaga agaatagctg ggagcctggg cagccaacgg gtcgacggag 2160
gcagaaattc ggcacatgga cagcccagcg ggggtcaaga catcgactcc gcgctagaaa 2220
cacagaacaa gcgacagaga gcggctcaac aacctgtcga aacggcgcaa ggatgcgcga 2280
caa 2283

<210> 19
<211> 1193
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7625645CB1

<400> 19
gaagaacttc atggtcggcg caggcagcat ccaggcggcg gtcgagaact tcgtgcgtga 60
ggtgaaggcg cgcaccttcc caagtccgga atactgctac tgacgcggcc tcaaggcgcg 120
tactcgtccg tcatggagat catccggacc atagccgagc tgcgggttcg tctcgcacgc 180
gaatcgaaca tcgcgctggt gcccaccatg ggcaacatcc acgacggcca cctgagcctc 240
gtgcgcattg gtctcaagcg cgccccgttc acggtgacca gcatcttcgt gaaccgcctc 300
cagttcatcc agggcgagga cttcgacaag tacccgcgca ccttcgagga ggattgcgag 360
aagctcgagg cggccggcaa cagcgtggtg ttcgcgccgg acgaaaccca gatgtatcct 420
gagcggcagg tgttcatcgt cgagccgccg cccatcgcga acaagctgga gggccgcttt 480
cgtccaggac atttccgcgg cgtgtccacg gtcgtcctca agctcttcaa catggtgcaa 540
ccgcaggtcg cggtgttcgg caagaaggac taccagcagc tttccatcgt gcgccatatg 600
gtggaccagc tcgcgttgcc cctgtccatc attcccgccg agacggtgcg cgccgaagat 660
ggcctggcac tttcgtcgcg caaccgctac ctgagccccg aggaacgcag cgaggccccg 720
cgcctcaacg aagcgctgcg ccaggtgaag caggccgtgg aatcgggtga tcgcaacttc 780
gaggcgctgg agttcgcagc cgatgcgctg cttgcgcgcc atcgttggca cgtggactat 840

gtcgtggtgc gcagccgccg cacgctgatg cagccggacc cggacgagcg cgaactggtg 900
gtgctgggtg cggcgcgtct tggacagacg cggctcatcg acaacctgga gatcctggcg 960
cccgcctgaa cgtcggcagg gcgctcatga aggcaagccc atgagccgtc ctgcaccgtt 1020
tccctttcg gccatcgtcg gtcaggacga gatgaagctc gccctgctga tcgcggccgt 1080
ggacccgacc gtcggtgggg tgctcgtgtt cggcgatcgc ggtaccggca agtccacagc 1140
cgtgcgggca ctcgcggctc tgctgccacc catgcgggcg gtcgcgggtt gcc 1193

<210> 20
<211> 2092
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5730123CB1

<400> 20
cggacgcgtg ggcggacgcg tgggggcag gtctggcggg actgcgcggc ggcaacagca 60
gacatgtcgg gggtccgggg cctgtcgcgg ctgctgagcg ctcggcgcct ggcgctggcc 120
aaggcgtggc caacagtgtt gcaaacagga acccgaggtt ttcacttcac tgttgatggg 180
aacaagaggg catctgctaa agtttcagat tccatttctg ctcagtatcc agtagtggat 240
catgaatttg atgcagtggt ggtaggcgct ggaggggcag gcttgcgagc tgcatttggc 300
ctttctgagg cagggtttaa tacagcatgt gttaccaagc tgtttcctac caggtcacac 360
actgttgcag cacagggagg aatcaatgct gctctgggga acatggagga ggacaactgg 420
aggtggcatt tctacgacac cgtgaagggc tccgactggc tgggggacca ggatgccatc 480
cactacatga cggagcaggc ccccgccgcc gtggtcgagc tagaaaatta tggcatgccg 540
tttagcagaa ctgaagatgg gaagatttat cagcgtgcat ttggtggaca gagcctcaag 600
tttggaaagg gcgggcaggc ccatcggtgc tgctgtgtgg ctgatcggac tggccactcg 660
ctattgcaca ccttatatgg aaggtctctg cgatatgata ccagctattt tgtggagtat 720
tttgccttgg atctcctgat ggagaacggg gagtgccgtg gtgtcatcgc actgtgcata 780
gaggacgggt ccatccatcg cataagagca aagaacactg ttgttgccac aggaggctac 840
gggcgcacct acttcagctg cacgtctgcc cacaccagca ctggcgacgg cacggccatg 900
atcaccaggg caggccttcc ttgccaggac ctagagtttg ttcagttcca ccccacaggc 960
atatatggtg ctggttgtct cattacggaa ggatgtcgtg gagagggagg cattctcatt 1020
aacagtcaag gcgaaaggtt tatggagcga tacgcccctg tcgcgaagga cctggcgtct 1080
agagatgtgg tgtctcggtc gatgactctg gagatccgag aaggaagagg ctgtggccct 1140
gagaaagatc acgtctacct gcagctgcac cacctacctc cagagcagct ggccacgcgc 1200
ctgcctggca tttcagagac agccatgatc ttcgctggcg tggacgtcac gaaggagccg 1260
atccctgtcc tccccaccgt gcattataac atgggcggca ttcccaccaa ctacaagggg 1320
caggtcctga ggcacgtgaa tggccaggat cagattgtgc ccggcctgta cgcctgtggg 1380
gaggccgcct gtgcctcggt acatggtgcc aaccgcctcg gggcaaactc gctcttggac 1440
ctggttgtct ttggtcgggc atgtgccctg agcatcgaag agtcatgcag gcctggagat 1500
aaagtccctc caattaaacc aaacgctggg gaagaatctg tcatgaatct tgacaaattg 1560
agatttgctg atggaagcat aagaacatcg gaactgcgac tcagcatgca gaaggtgcgg 1620
attgatgagt acgattactc caagcccatc caggggcaac agaagaagcc ctttgaggag 1680
cactggagga agcacaccct gtcctttgtg gacgttggca ctgggaaggt cactctggaa 1740
tatagacccg taatcgacaa aactttgaac gaggctgact gtgccaccat cccgccagcc 1800
attcgctcct actgatgaga caagatgtgg tgatgacaga atcagctttt gtaattatgt 1860
ataatagctc atgcatgtgt ccatgtcata actgtcttca tacgcttctg cactctgggg 1920
aagaaggagt acattgaagg gagattggca cctagtggct gggagcttgc caggaaccca 1980
gtggccaggg agcgtggcac ttacctttgt cccttgcttc attcttgtga gatgataaaa 2040
ctgggcacag ctcttaaata aaatataaat gaacaaaaaa aaaaaaacag gg 2092

<210> 21
<211> 4213
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7481031CB1

<400> 21

```
ggttgggtgg ggctttgggt cccatccatt cccgagcact gcggccctgg gcgaagggta 60
gggttggagg gatggggcag gattgactct gggctgcccg gcagaatcat gcggaagcag 120
ccgaccgtcg tcctgggtgt ggcgcacact gtggtgaaga ggatgtcgac cttcgtgcgg 180
caaatcgact ttgccctgga ctgggtggag gtggaggccg ggcgagcaat atacaggcag 240
ggggacaagt ccgactgcac gtacatcatg ctcagcggcc ggctgcgctc tgtgatccgg 300
aaggatgatg ggaagaagcg cctggccggg gagtacggcc gaggagacct cgtcggcgtg 360
cctggtggtt ccagctccat ccaccggatg aaggctcaca ctccaggccg gatgccggca 420
gggaggcctc agtgccagag ggcagggctc ccgcaggctg agggacacat cttaaaaaga 480
ggagttctcc aggccgggcg cggcggttca cgcctgtaat cccagcactt tgggaggccg 540
aggcggcgga ccatgaggtc aggagatcga gaccatcctg gtatgtggga ggcgggattg 600
ggcggggagg ggccagggct agtcagaggc cggcactcgc ttgcactcac ctgctggccc 660
ggcagatgca gaggcctggt ggccggtggg ctggtccaga gcgcgttcag gtgcgcaggg 720
cactgtgggg accctgctcc ctgcgtggcc ccgaatgggc tgcaccagac accagggagt 780
gcgtgcaaac aaggttgctt acggctgcct ccccagcaga tcacgggggc ggccgcccgg 840
cagctgctgg gcccctgctg aagaggagcc actccgtccc cgcgccttcc attcgcaaac 900
agatcttgga ggagctggag aagcccgggg caggtgaccc tgacccttcg gccccacaag 960
ggggcccagg cagtgccact tctgatctgg ggatggcatg tgaccgtgcc agggtcttcc 1020
tgcactcgga cgagcacccc gggagctccg tggccagcaa gtccaggaaa agcgtgatgg 1080
ttgcagagat accctccacg gtctcccagc actcagagag tcacacggat gagaccctgg 1140
ccagcaggaa gtcggatgcc atcttcagag ctgccaagaa ggacctgctc accctgatga 1200
agctggaaga ctcatctctg ttggatggcc gggtggcgct tctgcacgtt cctgcaggca 1260
cggtggtgtc aaggcaggga gaccaggacg ccagcatcct gttcgtggtc tcggggctgc 1320
tgcacgtgta ccagcggaag atcggcagcc aggaggacac ctgcttgttc ctcacgcgcc 1380
ccggggagat ggtgggccag ctggccgtgc tcaccgggga gcctctcatc ttcaccgtca 1440
aggccaacag ggactgcagc ttcctgtcca tctccaaggc ccacttctat gaaatcatgc 1500
ggaagcagcc gaccgtcgtc ctgggtgtgg cgcacactgt ggtgaagagg atgtcgtcct 1560
tcgtgcggca aatcgacttt gccctggact gggtggaggt ggaggccggg cgagcaatat 1620
acaggcaggg ggacaagtcc gactgcacgt acatcatgct cagcggccgg ctgcgctctg 1680
tgatccggaa ggatgatggg aagaagcgcc tggccgggga gtacggccga ggagacctcg 1740
tcggcgtggt ggagacactg acccaccagg cccgggcgac cacggtgcat gccgttcggg 1800
actcagaatt ggccaagctg ccggcaggag ccctcacgtc catcaagcgc aggtacccac 1860
aggtggtgac tcggctgatt catctcttgg gtgagaagat cctgggcagc ctccagcagg 1920
gacctgtgac aggccaccag cttgggctcc ccacggaggg cagcaagtgg gacttgggga 1980
acccggctgt caacctgtcc acggtggcag tgatgcccgt gtcagaggaa gtgcccctca 2040
ccgccttcgc cctggagctg gagcatgccc tcagcgccat cggcccgacc ctgctgctga 2100
ctagtgacaa cataaaacgg cgccttggct ccgctgccct ggacagtgtt cacgagtacc 2160
ggctgtccag ctggctgggg cagcaggagg acacccacag gatcgtgctc taccaggcag 2220
atggcacgct cacaccctgg acccagcgct gcgtgcgcca ggccgactgc atcctcatcg 2280
tgggcctggg tgaccaggag cccacagtgg gcgagctgga gcggatgctg gagagcacag 2340
ctgtgcgtgc ccagaagcag ctgatcctgc tgcacaggga ggagggcccg gcgccagcgc 2400
gcaccgtgga gtggctcaac atgcggagct ggtgctccgg ccacctgcac ctctgctgcc 2460
cgcgccgcgt cttctccagg aggagcctgc ccaagctggt ggagatgtac aagcatgtct 2520
tccagcggcc cccggaccga cactcagact tctcccgcct ggcgagggtg ctgacgggca 2580
acgccattgc cctggtgctt gggggagggg gagcaagagg ctgtgcccag gtgggcgttc 2640
tcaaggcctt ggcggagtgc ggcatccctg tggacatggt gggaggcacg tccatcgggg 2700
ccttcgtggg tgccctgtac tctgaggagc ggaactacag ccagatgcgg atccgggcca 2760
agcagtgggc cgagggcatg acgtccttga tgaaggccgc gctggacctc acctacccca 2820
tcacgtccat gttctccgga gccggcttca acagcagcat cttcagcgtc ttcaaggacc 2880
agcagatcga ggacctgtgg attccttatt tcgccatcac caccgacatc acagcctcgg 2940
ccatgcgggt ccacaccgac ggctccctgt ggcggtacgt gcgtgccagc atgtccctgt 3000
ccggttacat gccccctctc tgtgacccga aggacggaca cctgctgatg gacgggggct 3060
acatcaacaa cctcccagcg gatgtggccc ggtccatggg ggcaaaagtg gtgatcgcca 3120
ttgacgtggg cagccgagat gagacggacc tcaccaacta tggggatgcg ctgtctgggt 3180
ggtggctgct gtggaaacgc tggaacccct tggccacgaa agtcaaggtg ttgaacatgg 3240
cagagattca gacgcgcctg gcctacgtgt gttgcgtgcg gcagctggag gtggtgaaga 3300
gcagtgacta ctgcgagtac ctgcgccccc ccatcgacag ctacagcacc ctggacttcg 3360
gcaagttcaa cgagatctgc gaagtgggct accagcacgg gcgcacggtg tttgacatct 3420
ggggccgcag cggcgtgctg gagaagatgc tgcgcgacca gcaggggccg agcaagaagc 3480
ccgcgagtgc ggtcctcacc tgtcccaacg cctccttcac ggaccttgcc gaaattgtgt 3540
ctcgcattga gcccgccaag cccgccatgg tggatgacga atctgactac cagacggagt 3600
acgaggagga gctgctggac gtccccaggg atgcatacgc agacttccag agcacctcag 3660
cccagcaggg ctcagacttg gaggacgagt cctcactgcg gcatcgacac cccagtctgg 3720
ctttcccaaa actgtctgag ggctcctctg accaggacgg gtagaggcct ctgctaaaga 3780
```

gcccggatgc agcgtcttcc gtgggactgt ccccaaggct gaggctcctg ccaagtccta 3840
ggggcctctg tacctgccct gctggaagcc ctgacttccc cggggcccca ggctgtgtta 3900
gggttctctg ggcctcttct ttgtaccagc agccctgcat acagggccct gtgagccccc 3960
ctgcagtcct gtgaggcccc tgaagctctg tgaggcccct gaagctctgt gaacccccctg 4020
cagccctgtg aggccccccg aagccctgtg aggccccccg aagccctgtg aaccacctgc 4080
tgccctgtga ggcccccaaa gccctgtgaa ctgcctgctg tcctgtgaac tgcctgctgc 4140
cctgtgaggt gtgggagccc tgatgctgcc gtgtgatgtt tcaataaagg tggatctcac 4200
tgttaaaaca aaa 4213

<210> 22
<211> 1784
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7491216CB1

<400> 22
gcttcctctc agcccgctct gagctggaag cagcatgtgg gacctggccc tgatcttcct 60
cgcagcagcc tgcgtgttct cactaggggt cactctgtgg gtcatttgca gccatttttt 120
cactgtgcac atccctgcag cggttggcca ccctgtgaaa ctgagagtcc tccattgcat 180
cttccagctg ctgttgactt gggggatgat atttgagaag ctcagaatct gttctatgcc 240
ccaatttttc tgtttcatgc aagatctgcc tccgctaaag tatgaccccg atgttgtggt 300
cacggatttc cgctttggga caatccctgt gaagctgtac caacccaagg catccacctg 360
caccctgaag cctggcatcg tgtactacca cggtggcggg ggcgtcatgg ggagtttgaa 420
aacccaccat ggcatatgct ctcgtttgtg caaggagagt gactccgtgg ttctggcagt 480
tggttaccgc aagttaccta agcataagtt tccagtgcca gtaagagact gcttggtggc 540
caccatccac ttcctgaagt ccctggatgc atatggagtg gatccagccc gggttgtggt 600
ctgcggtgac agtttcggag gggcaatagc cgcagtggtt tgtcaacaac ttgtggacag 660
gccagatctg ccccggatcc gggctcagat cctgatctat gccattctcc aagccctgga 720
tttacaaacc ccttcgtttc aacagaggaa aaacatccca ctgctcacct ggagtttcat 780
ctgctactgt ttttttcaaa acctggattt cagctcctcc tggcaagagg tcatcatgaa 840
aggtgcccat ttgcctgctg aagtctggga aaagtacaga aagtggttgg gcccagaaaa 900
catccctgag aggtttaagg agaggggtta ccaactgaag ccccatgagc ccatgaatga 960
agctgcttac ttggaagtaa gtgttgtcct ggatgtgatg tgctcgcccc tgattgcaga 1020
agatgacata gtgtctcagc tcccggaaac ctgcatcgtg agctgtgagt atgatgctct 1080
ccgggacaat tcactgttgt acaagaaaag gctggaagac ctgggagtgc ccgtgacctg 1140
gcaccatatg gaggatggtt tccatggagt gctcaggacc attgacatga gcttcttgca 1200
ctttccctgc tccatgagaa ttctgagtgc attagttcaa tttgtaaagg gactgtgacc 1260
atctttcttc tctgctggta ctgcggtgtg gattccactg gcatccagcc tcccacaggg 1320
ctctctgttg ctgatttagg tggtgcatag tggggctagg gaggaggtag aggttgctgt 1380
caccttttctg gtccaggttc tagaaccaca caatgcatgc tcctgatgtc cagaggacgt 1440
ggtagaaaag acaggtttgg aggtgggagt gtggctgtct ctattctctg ttgggaaaac 1500
ctgggctgac aatattcagt ggccatttgt gggagtgaat cagccggtaa gagctgttct 1560
cagcctctct aaggggcagt tcaggctccc agattgatcc agactgtgtg tgactttcgt 1620
ccatttgact tgactttgga atagcacaag ggcatcacgt acttcacgag gctttcccaa 1680
tgtggctcag aggcaggagc tctgatgctc taggctgctg tgaggtggtg gtggtggtgg 1740
agaaactggc ttcacccacc tactcttctg tgaacagtag tgac 1784

<210> 23
<211> 2141
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 71624817CB1

<400> 23
ctcttttgtg ctgattcctg aggactagga aggtgccccg aaaagaattc agagtgagta 60
cagtgaaaca gaaacctgct cagcttctaa gtgcaggaag gacttcacag ggaggcatga 120
tcagaacttg aaaaaaggac ctgacaatgt ccctgtattg tggaatagct tgcaggagaa 180

aatttttttg gtgctatagg ctgctgtcaa cctatgttac taagacacgg tatttatttg 240
aactgaagga agatgatgat gcatgtaaaa aagcccagca aacaggagcg ttttacctct 300
ttcatagtct ggctcctctg cttcagactt cagcacatca atacctggcc ccccggcaca 360
gcctgttaga gttggaaagg ctccctgggta aatttggaca ggatgcacaa agaatagaag 420
attctgtgct gattggatgc tctgagcagc aggaagcatg gtttgctctg gatctaggtc 480
tggatagctc cttttccata agtgcctcct tacacaaacc tgaaatggag acagagctca 540
aggggtcttt cattgagctg agaaaggcac tctttcaact caatgcaagg gatgcctcct 600
tgctgtccac ggctcaagct cttctccgct ggcatgatgc tcatcagttc tgcagcagaa 660
gtgggcagcc caccaagaag aacgtggctg gcagcaagcg tgtgtgccct tccaataata 720
taatctatta tccacagatg gctcctgtgg cgatcacgct ggtgtcagat gggacccgat 780
gcctgcttgc ccgccaaagc tcctttccca agggaatgta ttctgccttg gcaggttttt 840
gtgatatagg tgaaagtgtg gaagagacca tccgccgaga agttgcagaa gaggtgggat 900
tggaggtgga aagcctgcag tactatgcat cccagcattg gcccttccct agtggctcac 960
tcatgattgc ttgccatgca actgtgaaac cagggcagac agaaatccag gtgaacttga 1020
gagaattaga gacagctgcc tggttcagtc atgatgaggt agccacagcc ctgaagagaa 1080
agggccccta tactcagcaa cagaatggga ctttcccatt ctggctgccc cctaagttag 1140
ccatctccca ccaactgatt aaggagtggg tggaaaaaca gacctgttct tccctgcctg 1200
cttagcccgg atcaagtcac ttagatcgct ccttggtatt cctgagggac aaactagaga 1260
tcagttgaca aaggagaagt gacaaaagat aagctgcaga aggacctcag aagggcagag 1320
caaagggtga gcctacagta agacacttct atcagcagtg ttaatggaag aaattcctac 1380
caatgggcaa tcaaaaaagc cagtgtgaga agaaaactga tgagctgtca actgtcaaaa 1440
atcaggggga agggggaagc attagtttgg atgtaggccc ttgttcagtc attttctcta 1500
aggccttgga agcaaacatc ttctggcata tggcttgtta atgttgtgtt tacactaact 1560
gccaaaatgt gcctgttgta agtttggtta aaacttttat cttagtattg aaaatatata 1620
gcaagtttta agtcattact gagttacttg ttacccttac agaattaaga aaaatagcac 1680
agtatgacat tttaaaattc atgtacaggt gcattctagt tacatgaaca tgctagttaa 1740
ataaaagtca caattaagtc attatgactc agagtacttt ataataaatc agatgccctg 1800
aggctctgtt gaggaagtag aatcccaatg ggactccaga cagtaaatct ttttcgctgt 1860
tttcagctct gtcacatgct cagtttgtga cccagggaaa agactattgc tttgccatgc 1920
ctgttttcct aaataaaaca gttgctggca tttgcacatc cacaacgtgt tgattaacaa 1980
agggattggt ggagataaac agatgccaaa cctgacctat ttcttaaact ttatggaata 2040
aactaaattt aggatttctc atcattcata tatgatgcca taagggagaa gagtttattt 2100
ggggaaaata aaagaaattt ccacaaaaaa aaaaaaaaaa a 2141

<210> 24
<211> 2518
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 6945964CB1

<400> 24
tgaagacagc ttccatcacg ttactaattc gattcgctcg agccactgag gcttttacaa 60
gaggaaagga gttgctctta gcagatgaca gacacttctc aaaagacagc ttttcttcct 120
ggagaacaga ctttttcagc aggattttcc tttcagtgaa acataatttg acttgaaagg 180
aacccaggga aaagtgtcca ggtgtgagca tgagcgggta gaggtgtgcc cttgtttgct 240
tcaggctgtc tgcttttcgc ccctgactgt tttttctgtt tctggccatg gaggaagaga 300
aagatgacag cccacaggct gacttctgcc tgggcaccgc cctgcactct tggggactgt 360
ggttcacgga ggaaggttca ccgtccacca tgctgacggg gattgcagtt ggagccctcc 420
tggccctggc cttggttggt gtcctcatcc ttttcatgtt cagaaggctt agacaatttc 480
gacaagcaca gcccactcct cagtaccggt tccggaagag agacaaagtg atgttttacg 540
gccggaagat catgaggaag gtgaccacac tccccaacac ccttgtggag aacactgccc 600
tgccccggca gcgggccagg aagaggacca aggtgctgtc tttggccaag aggattctgc 660
gtttcaagaa ggaatacccg gccctgcagc ccaaggagcc cccgccctcc ctgctggagg 720
ccgacctcac ggagtttgac gtgaagaatt ctcacctgcc atcggaagtt ctgtacatgc 780
tgaaaaacgt tcgggtcctg ggccactttg agaagccgct gttcctggag ctttgcaaac 840
acatcgtctt tgtgcagctg caggaagggg agcacgtctt ccagcccagg gagccggacc 900
ccagcatctg tgtggtgcag gacgggcggc tggaggtctg catccaggac actgacggca 960
ccgaggtggt ggtgaaagag gttctggcgg gagacagcgt ccacagcctg ctcagcatcc 1020
tggacatcat caccggccat gctgcacctt acaaaacggt ctccgtccgc gcggccatcc 1080
cgtccaccat cctccggctt ccagctgcgg cttttcatgg agtttttgag aaatatccgg 1140

aaactctggt gagggtggtg cagatcatca tggtgcggct gcagagggtg acctttctgg 1200
ctctgcacaa ctacctcggc ctgaccacag agctcttcaa cgctgagagc caggccatcc 1260
ctctcgtgtc tgtagccagt gtggctgccg ggaaggccaa gaagcaggtg ttctatggcg 1320
aagaagagcg gcttaaaaag ccaccgcggc tccaggagtc ctgtgactca ggtactgtcc 1380
tgcaccaagg agggcaatgt ccagccccag agtgggggat ggagtggaca gagtggagcg 1440
agtgagtggg ggatggagtg gacagagtgg agtgagtggg ggatggagtg agcagagttc 1500
gagaccagcc cgggcaacat agtgagaccc tgtctctatg aataaaattt gaaaaatgat 1560
ccaggcatgt tagcacacac ctgtcgtccc agctatgcag agggttgagg tgggaggatc 1620
gtttgggcct cggagttcga ggttacagtg agctatgatc gcgccaccgc agtccagcct 1680
ggtcaacaga gcgagaccct gcctcaaaac acacacacaa aaaggtgttt ttctctgtct 1740
ttagttttca gaaatgtgat ttggaagtgt ctgagcatga atactttgga tctgtcctgt 1800
ttggggtttc ttcgtcttgg acccgtaggt ttatagcttt taacacattt tggaagtgtt 1860
cagtctatat gtcttcaaac gattcttcca ctcgtgtgcc cctccacctc ttctggaaat 1920
aaagtggcac aagtgttaga cctcgcgtta tgtccccaag gtctctgggg cactgctcac 1980
tttccttaaa ccttttttgtt ctctctctga tgttaagatg gaattgtttc ggtttgatct 2040
atcttcacgg atctttggct tgtttcgtca tctcgattct agctttgagg cctatcagag 2100
agttttaaaa aaatgttctg tcatcgtgtt ttggttcttc tgtgtgcctc ctgtttcttt 2160
gctaggacgt tctctctttg tctttgtctc aggagtgttc aggattgctg gttctagctg 2220
gtaatgataa cagctttatc gtcctggaga aataattccg tctgtgacat gcggtcgttt 2280
ttatctgtgg attgttttta ctcctctggt tcagatcctc cttgctcttg gcatactggg 2340
tcaccatgta gtcgtcgtac tgttcactgg aattatccgg atactgtcgt catatctggt 2400
gttcgaatat tgtgacgacg gataaccact gtgcactcat ggacacaagt ctaagtaccc 2460
actgagttac ccaagccagt gtgctcgagt ccagaacaca cctgacgacc gcccagct 2518